

Supplementary Materials for “Mendelian randomization
accounting for complex correlated horizontal pleiotropy
while elucidating shared genetic etiology”

Qing Cheng^{1,2}, Xiao Zhang², Lin S. Chen^{3*}, Jin Liu^{2*}

¹ Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan, China

²Centre for Quantitative Medicine, Health Services & Systems Research,
Duke-NUS Medical School

³Department of Public Health Sciences, The University of Chicago

*Corresponding authors. E-mails: lchen@health.bsd.uchicago.edu;
jin.liu@duke-nus.edu.sg;

Supplementary Notes

Supplementary Note 1. The Gibbs sampler algorithms for MR-CUE estimation	3
1.1 For independent SNPs	3
1.2 For correlated SNPs	5
Supplementary Note 2. Pseudo-code	8
Supplementary Note 3. Additional simulation results	13
3.1 The identifiability of causal effect	13
3.2 Simulation results for different p and $\rho_{\alpha\gamma}$	15
3.3 Generation of summary statistics for multiple confounding pathways	25
3.4 Simulation results for non-linear confounders	26
3.5 Simulation results for binary outcome	27
3.6 Simulation results for different proportions of IVs with CHP effects	27
3.7 Simulation results for sparse θ in reverse causation analysis	28
Supplementary Note 4. Sensitive Analysis	29
4.1 Negative and Positive Controls	29
Supplementary Note 5. Real Data Analysis	32
Supplementary References	35

Supplementary Note 1

The Gibbs sampler algorithms for MR-CUE estimation

In this section, we present the details on deriving Gibbs sampler for MR-CUE.

1.1 For independent SNPs

We propose the following hierarchical Bayesian model for independent SNPs.

$$\begin{aligned}
\begin{pmatrix} \hat{\gamma}_k \\ \hat{\Gamma}_k \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} \gamma_k \\ \Gamma_k \end{pmatrix}, \begin{pmatrix} \hat{s}_{\gamma_k}^2, & \hat{\rho}_e \hat{s}_{\gamma_k} \hat{s}_{\Gamma_k} \\ \hat{\rho}_e \hat{s}_{\gamma_k} \hat{s}_{\Gamma_k}, & \hat{s}_{\Gamma_k}^2 \end{pmatrix}\right), \\
\gamma_k | \sigma_{\gamma}^2 &\stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_{\gamma}^2), \quad \Gamma_k | \beta_1, \beta_2, \gamma_k, \eta_k, \tau_1^2, \tau_2^2, \xi^2 \stackrel{i.i.d.}{\sim} \left\{ \mathcal{N}(\beta_1 \gamma_k, \xi^2 \tau_1^2) \right\}^{(1-\eta_k)} \left\{ \mathcal{N}(\beta_2 \gamma_k, \tau_2^2) \right\}^{\eta_k}, \\
\sigma_{\gamma}^2 &\sim \mathcal{IG}(a_{\gamma}, b_{\gamma}), \quad \tau_1^2 \sim \mathcal{IG}(a_{\tau_1}, b_{\tau_1}), \quad \tau_2^2 \sim \mathcal{IG}(a_{\tau_2}, b_{\tau_2}), \\
\Pr(\xi^2) &\propto \frac{1}{\xi^2}, \quad \eta_k | \omega \stackrel{i.i.d.}{\sim} \omega^{\eta_k} (1-\omega)^{1-\eta_k}, \quad \omega \sim Beta(a, b).
\end{aligned} \tag{1}$$

Based on the Bayesian model (1), the joint posterior density of all the latent variables is

$$\begin{aligned}
&\Pr(\boldsymbol{\Gamma}, \boldsymbol{\gamma}, \sigma_{\gamma}^2, \tau_1^2, \tau_2^2, \xi^2, \boldsymbol{\eta}, \omega | \hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\gamma}}) \\
&\propto \Pr(\hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\gamma}} | \boldsymbol{\Gamma}, \boldsymbol{\gamma}) \Pr(\boldsymbol{\Gamma} | \beta_1, \beta_2, \boldsymbol{\gamma}, \boldsymbol{\eta}, \tau_1^2, \tau_2^2, \xi^2) \Pr(\boldsymbol{\gamma} | \sigma_{\gamma}^2) \Pr(\boldsymbol{\eta} | \omega) \Pr(\xi^2) \Pr(\sigma_{\gamma}) \Pr(\tau_1^2) \Pr(\tau_2^2) \Pr(\omega) \\
&= \prod_{k=1}^p \left\{ \Pr(\hat{\Gamma}_k, \hat{\gamma}_k | \Gamma_k, \gamma_k) \Pr(\Gamma_k | \beta_1, \beta_2, \gamma_k, \eta_k, \tau_1^2, \tau_2^2, \xi^2) \Pr(\gamma_k | \sigma_{\gamma}^2) \Pr(\eta_k | \omega) \right\} \\
&\quad \Pr(\xi^2) \Pr(\sigma_{\gamma}) \Pr(\tau_1^2) \Pr(\tau_2^2) \Pr(\omega)
\end{aligned}$$

The factors in the joint posterior density that involve Γ_k form a mixture of two normal distributions. The conditional posterior distribution of each Γ_k given the other parameters in the model is

$$\Gamma_k | \hat{\Gamma}_k, \gamma_k, \eta_k, \beta_1, \beta_2, \tau_1^2, \tau_2^2, \xi^2 \sim (1 - \eta_k) \mathcal{N}(\tilde{\mu}_{k0}, \tilde{\sigma}_{k0}^2) + \eta_k \mathcal{N}(\tilde{\mu}_{k1}, \tilde{\sigma}_{k1}^2), \tag{2}$$

where $(\tilde{\mu}_{k0}, \tilde{\sigma}_{k0}^2)$ and $(\tilde{\mu}_{k1}, \tilde{\sigma}_{k1}^2)$ satisfy

$$\begin{cases} -\frac{1}{2\tilde{\sigma}_{k0}^2} = -\frac{1}{2} \left(\frac{1}{(1-\tilde{\rho}_e^2)\tilde{s}_{\Gamma_k}^2} + \frac{1}{\xi^2\tau_1^2} \right), \\ \frac{\tilde{\mu}_{k0}}{\tilde{\sigma}_{k0}^2} = \frac{\hat{\Gamma}_k}{(1-\tilde{\rho}_e^2)\tilde{s}_{\gamma_k}^2} + \frac{\tilde{\rho}_e(\gamma_k - \hat{\gamma}_k)}{(1-\tilde{\rho}_e^2)\tilde{s}_{\gamma_k}\tilde{s}_{\Gamma_k}} + \frac{\beta_1\gamma_k}{\xi^2\tau_1^2}, \end{cases} \quad \text{and} \quad \begin{cases} -\frac{1}{2\tilde{\sigma}_{k1}^2} = -\frac{1}{2} \left(\frac{1}{(1-\tilde{\rho}_e^2)\tilde{s}_{\Gamma_k}^2} + \frac{1}{\tau_2^2} \right), \\ \frac{\tilde{\mu}_{k1}}{\tilde{\sigma}_{k1}^2} = \frac{\hat{\Gamma}_k}{(1-\tilde{\rho}_e^2)\tilde{s}_{\gamma_k}^2} + \frac{\tilde{\rho}_e(\gamma_k - \hat{\gamma}_k)}{(1-\tilde{\rho}_e^2)\tilde{s}_{\gamma_k}\tilde{s}_{\Gamma_k}} + \frac{\beta_2\gamma_k}{\tau_2^2}. \end{cases}$$

Conditioning on the data $(\hat{\Gamma}, \hat{\gamma})$ and other parameters, the conditional distribution for each element γ_k comes from a mixture of two normal distributions, i.e.,

$$\gamma_k | \hat{\gamma}_k, \Gamma_k, \eta_k, \beta_1, \beta_2, \tau_1^2, \tau_2^2, \sigma_{\gamma}^2, \xi^2 \sim (1 - \eta_k) \mathcal{N}(\mu_{k0}, \sigma_{k0}^2) + \eta_k \mathcal{N}(\mu_{k1}, \sigma_{k1}^2), \quad (3)$$

where $(\mu_{k0}, \sigma_{k0}^2)$ and $(\mu_{k1}, \sigma_{k1}^2)$ satisfy

$$\begin{cases} -\frac{1}{2\sigma_{k0}^2} = -\frac{1}{2} \left(\frac{1}{(1-\tilde{\rho}_e^2)\tilde{s}_{\gamma_k}^2} + \beta_1^2 \frac{1}{\xi^2\tau_1^2} + \frac{1}{\sigma_{\gamma}^2} \right), \\ \frac{\mu_{k0}}{\sigma_{k0}^2} = \frac{\hat{\gamma}_k}{(1-\tilde{\rho}_e^2)\tilde{s}_{\gamma_k}^2} + \frac{\tilde{\rho}_e(\Gamma_k - \hat{\Gamma}_k)}{(1-\tilde{\rho}_e^2)\tilde{s}_{\gamma_k}\tilde{s}_{\Gamma_k}} + \frac{\beta_1\Gamma_k}{\xi^2\tau_1^2}, \end{cases} \quad \text{and} \quad \begin{cases} -\frac{1}{2\sigma_{k1}^2} = -\frac{1}{2} \left(\frac{1}{(1-\tilde{\rho}_e^2)\tilde{s}_{\gamma_k}^2} + \beta_2^2 \frac{1}{\tau_2^2} + \frac{1}{\sigma_{\gamma}^2} \right), \\ \frac{\mu_{k1}}{\sigma_{k1}^2} = \frac{\hat{\gamma}_k}{(1-\tilde{\rho}_e^2)\tilde{s}_{\gamma_k}^2} + \frac{\tilde{\rho}_e(\Gamma_k - \hat{\Gamma}_k)}{(1-\tilde{\rho}_e^2)\tilde{s}_{\gamma_k}\tilde{s}_{\Gamma_k}} + \frac{\beta_2\Gamma_k}{\tau_2^2}. \end{cases}$$

The conditional posterior distributions of β_1 and β_2 are both from normal, i.e.,

$$\beta_1 | \Gamma, \gamma, \eta, \tau_1^2, \xi^2 \sim \mathcal{N}(\mu_{\beta0}, \sigma_{\beta0}^2), \quad \text{and} \quad \beta_2 | \Gamma, \gamma, \eta, \tau_2^2 \sim \mathcal{N}(\mu_{\beta1}, \sigma_{\beta1}^2), \quad (4)$$

where

$$\begin{cases} -\frac{1}{2\sigma_{\beta0}^2} = -\frac{\sum_{k=1}^p (1-\eta_k)\gamma_k^2}{2\xi^2\tau_1^2}, \\ \frac{\mu_{\beta0}}{\sigma_{\beta0}^2} = \frac{\sum_{k=1}^p (1-\eta_k)\Gamma_k\gamma_k}{\xi^2\tau_1^2}, \end{cases} \quad \text{and} \quad \begin{cases} -\frac{1}{2\sigma_{\beta1}^2} = -\frac{\sum_{k=1}^p \eta_k\gamma_k^2}{2\tau_2^2}, \\ \frac{\mu_{\beta1}}{\sigma_{\beta1}^2} = \frac{\sum_{k=1}^p \eta_k\Gamma_k\gamma_k}{\tau_2^2}, \end{cases}$$

Conditioning on the data and the other parameters in the model, the conditional posterior distributions of $\sigma_{\gamma}^2, \tau_1^2, \tau_2^2, \xi^2$ are all inverse-gamma, i.e.,

$$\begin{aligned} \sigma_{\gamma}^2 | \gamma, a_{\gamma}, b_{\gamma} &\sim \mathcal{IG}(\tilde{a}_{\gamma}, \tilde{b}_{\gamma}), & \tau_1^2 | \Gamma, \gamma, \xi^2, a_{\tau1}, b_{\tau1} &\sim \mathcal{IG}(\tilde{a}_{\tau1}, \tilde{b}_{\tau1}), \\ \tau_2^2 | \Gamma, \gamma, a_{\tau2}, b_{\tau2} &\sim \mathcal{IG}(\tilde{a}_{\tau2}, \tilde{b}_{\tau2}), & \xi^2 | \Gamma, \gamma, \beta_1, \eta &\sim \mathcal{IG}(a_{\xi}, b_{\xi}). \end{aligned} \quad (5)$$

where $\tilde{a}_{\gamma} = a_{\gamma} + \frac{p}{2}$, $\tilde{b}_{\gamma} = b_{\gamma} + \frac{\gamma^T \gamma}{2}$, $\tilde{a}_{\tau1} = a_{\tau1} + \frac{\sum_{k=1}^p (1-\eta_k)}{2}$, $\tilde{b}_{\tau1} = b_{\tau1} + \frac{\sum_{k=1}^p (1-\eta_k)(\Gamma_k - \beta_1\gamma_k)^2}{2\xi^2}$, $\tilde{a}_{\tau2} = a_{\tau2} + \frac{\sum_{k=1}^p \eta_k}{2}$, $\tilde{b}_{\tau2} = b_{\tau2} + \frac{\sum_{k=1}^p \eta_k(\Gamma_k - \beta_2\gamma_k)^2}{2}$, $a_{\xi} = \frac{1}{2} \sum_{k=1}^p (1 - \eta_k)$ and $b_{\xi} = \frac{1}{2\tau_1^2} \sum_{k=1}^p (1 - \eta_k)(\Gamma_k - \beta_1\gamma_k)^2$.

The conditional posterior of ω is also a Beta distribution:

$$\omega|\boldsymbol{\eta}, a, b \sim \text{Beta}\left(a + \sum_{k=1}^p \eta_k, b + \sum_{k=1}^p (1 - \eta_k)\right). \quad (6)$$

The conditional probability of $\boldsymbol{\eta}$ given $\mathbf{\Gamma}$ can be found using Bayes' theorem

$$\Pr(\eta_k = 1 | \Gamma_k) = \frac{\Pr(\eta_k = 1) \Pr(\Gamma_k | \eta_k = 1)}{\Pr(\eta_k = 0) \Pr(\Gamma_k | \eta_k = 0) + \Pr(\eta_k = 1) \Pr(\Gamma_k | \eta_k = 1)} \quad (7)$$

Algorithm 1 The Gibbs sampler algorithm for independent SNPs

```

1: Initialize:

2: repeat
3:   for  $k = 1$  to  $p$  do
4:     Sample  $\Gamma_k$  using (2).
5:     Sample  $\gamma_k$  using (3).
6:     Sample  $\eta_k$  using (7).
7:   end for
8:   Sample  $\beta_1$  and  $\beta_2$  using (4).
9:   Sample  $\sigma_\gamma^2, \tau_1^2, \tau_2^2, \xi^2$  using (5).
10:  Sample  $\omega$  using (6).
11: until Maximum iteration reached

```

1.2 For correlated SNPs

For correlated SNPs, we consider the following hierarchical Bayesian model for the l -block,

$l = 1, \dots, L$:

$$\begin{aligned}
\begin{pmatrix} \hat{\gamma}_l \\ \hat{\mathbf{\Gamma}}_l \end{pmatrix} &\sim \mathcal{N} \left(\begin{pmatrix} \hat{\mathbf{S}}_{\gamma_l} \hat{\mathbf{R}}^{(l)} \hat{\mathbf{S}}_{\gamma_l}^{-1} \gamma_l \\ \hat{\mathbf{S}}_{\mathbf{\Gamma}_l} \hat{\mathbf{R}}^{(l)} \hat{\mathbf{S}}_{\mathbf{\Gamma}_l}^{-1} \mathbf{\Gamma}_l \end{pmatrix}, \begin{pmatrix} \hat{\mathbf{S}}_{\gamma_l}, & \mathbf{0} \\ \mathbf{0}, & \hat{\mathbf{S}}_{\mathbf{\Gamma}_l} \end{pmatrix} (\mathbf{R}_e \otimes \hat{\mathbf{R}}^{(l)}) \begin{pmatrix} \hat{\mathbf{S}}_{\gamma_l}, & \mathbf{0} \\ \mathbf{0}, & \hat{\mathbf{S}}_{\mathbf{\Gamma}_l} \end{pmatrix} \right), \quad (8) \\
\Gamma_{lk} | \beta_1, \beta_2, \gamma_{lk}, \eta_l, \tau_1^2, \tau_2^2, \xi^2 &\stackrel{i.i.d.}{\sim} \left\{ \mathcal{N}(\beta_1 \gamma_{lk}, \xi^2 \tau_1^2) \right\}^{(1-\eta_l)} \left\{ \mathcal{N}(\beta_2 \gamma_{lk}, \tau_2^2) \right\}^{\eta_l}, \\
\gamma_{lk} | \sigma_\gamma^2 &\stackrel{iid}{\sim} \mathcal{N}(0, \sigma_\gamma^2), \quad \eta_l | \omega \stackrel{iid}{\sim} \omega^{\eta_l} (1 - \omega)^{1 - \eta_l}, \\
\sigma_\gamma^2 &\sim \mathcal{IG}(a_\gamma, b_\gamma), \quad \tau_1^2 \sim \mathcal{IG}(a_{\tau 1}, b_{\tau 1}), \quad \tau_2^2 \sim \mathcal{IG}(a_{\tau 2}, b_{\tau 2}), \\
\Pr(\xi^2) &\propto \frac{1}{\xi^2}, \quad \omega \sim \text{Beta}(a, b).
\end{aligned}$$

Based on model (8), the joint posterior density of all the parameters can be written as follows,

$$\begin{aligned}
& \Pr(\boldsymbol{\Gamma}, \boldsymbol{\eta}, \beta_1, \beta_2, \tau_1^2, \tau_2^2, \xi^2 | \widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\gamma}}) \\
& \propto \Pr(\widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{\gamma}} | \boldsymbol{\Gamma}, \boldsymbol{\gamma}) \Pr(\boldsymbol{\Gamma} | \beta_1, \beta_2, \boldsymbol{\gamma}, \boldsymbol{\eta}, \tau_1^2, \tau_2^2, \xi^2) \Pr(\boldsymbol{\gamma} | \sigma_{\boldsymbol{\gamma}}^2) \Pr(\boldsymbol{\eta} | \omega) \Pr(\xi^2) \Pr(\sigma_{\boldsymbol{\gamma}}) \Pr(\tau_1^2) \Pr(\tau_2^2) \Pr(\omega) \\
& = \prod_{l=1}^L \left[\Pr(\widehat{\boldsymbol{\Gamma}}_l, \widehat{\boldsymbol{\gamma}}_l | \boldsymbol{\Gamma}_l, \boldsymbol{\gamma}_l) \Pr(\eta_l | \omega) \Pr \left\{ \sum_{k=1}^{p_l} \Pr(\Gamma_{lk} | \beta_1, \beta_2, \gamma_{lk}, \eta_l, \tau_1^2, \tau_2^2, \xi^2) \Pr(\gamma_{lk} | \sigma_{\boldsymbol{\gamma}}^2) \right\} \right] \\
& \quad \Pr(\xi^2) \Pr(\sigma_{\boldsymbol{\gamma}}) \Pr(\tau_1^2) \Pr(\tau_2^2) \Pr(\omega)
\end{aligned}$$

For the k -th element in l -block, the conditional distribution of Γ_{lk} comes from a mixture distribution, i.e.,

$$\Gamma_{lk} | \widehat{\boldsymbol{\Gamma}}_l, \widehat{\boldsymbol{\gamma}}_l, \beta_1, \beta_2, \tau_1^2, \tau_2^2, \xi^2 \sim (1 - \eta_l) \mathcal{N}(\tilde{\mu}_{lk0}, \tilde{\sigma}_{lk0}^2) + \eta_l \mathcal{N}(\tilde{\mu}_{lk1}, \tilde{\sigma}_{lk1}^2), \quad (9)$$

where

$$\begin{aligned}
-\frac{1}{2\tilde{\sigma}_{lk0}^2} &= -\frac{1}{2} \left(\frac{1}{1 - \tilde{\rho}_e^2} \frac{\widehat{\mathbf{R}}_{kk}^{(l)}}{\widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_{lk}}^2} + \frac{1}{\xi^2 \tau_1^2} \right) \\
\frac{\tilde{\mu}_{lk0}}{\tilde{\sigma}_{lk0}} &= \frac{1}{1 - \tilde{\rho}_e^2} \left\{ \frac{\widehat{\Gamma}_{lk}}{\widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_{lk}}^2} - \sum_{j \neq k} \left(\frac{\widehat{\mathbf{R}}_{jk}^{(l)} \Gamma_{lj}}{\widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_{lj}}} \right) \frac{1}{\widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_{lk}}} - \tilde{\rho}_e \frac{\widehat{\gamma}_{lk}}{\widehat{\mathbf{s}}_{\boldsymbol{\gamma}_{lk}} \widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_{lk}}} + \left(\sum_{i=1}^p \frac{\widehat{\mathbf{R}}_{jk}^{(l)} \gamma_{li}}{\widehat{\mathbf{s}}_{\boldsymbol{\gamma}_{li}}} \right) \frac{\widehat{\rho}_e}{\widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_{lk}}} \right\} + \frac{\beta_1 \gamma_{lk}}{\xi^2 \tau_1^2} \\
-\frac{1}{2\tilde{\sigma}_{lk1}^2} &= -\frac{1}{2} \left(\frac{1}{1 - \tilde{\rho}_e^2} \frac{\widehat{\mathbf{R}}_{kk}^{(l)}}{\widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_{lk}}^2} + \frac{1}{\tau_2^2} \right) \\
\frac{\tilde{\mu}_{lk1}}{\tilde{\sigma}_{lk1}} &= \frac{1}{1 - \tilde{\rho}_e^2} \left\{ \frac{\widehat{\Gamma}_{lk}}{\widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_{lk}}^2} - \sum_{j \neq k} \left(\frac{\widehat{\mathbf{R}}_{jk}^{(l)} \Gamma_{lj}}{\widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_{lj}}} \right) \frac{1}{\widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_{lk}}} - \tilde{\rho}_e \frac{\widehat{\gamma}_{lk}}{\widehat{\mathbf{s}}_{\boldsymbol{\gamma}_{lk}} \widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_{lk}}} + \left(\sum_{i=1}^p \frac{\widehat{\mathbf{R}}_{jk}^{(l)} \gamma_{li}}{\widehat{\mathbf{s}}_{\boldsymbol{\gamma}_{li}}} \right) \frac{\widehat{\rho}_e}{\widehat{\mathbf{s}}_{\boldsymbol{\Gamma}_{lk}}} \right\} + \frac{\beta_2 \gamma_{lk}}{\tau_2^2}
\end{aligned}$$

For the k -th element in the l -block, the conditional distribution of γ_{lk} also comes from a mixture of two normal distributions, i.e.,

$$\gamma_{lk} | \widehat{\boldsymbol{\gamma}}_l, \boldsymbol{\Gamma}_l, \eta_l, \beta_1, \beta_2, \tau_1^2, \tau_2^2, \sigma_{\boldsymbol{\gamma}}^2, \xi^2 \sim (1 - \eta_l) \mathcal{N}(\mu_{lk0}, \sigma_{lk0}^2) + \eta_l \mathcal{N}(\mu_{lk1}, \sigma_{lk1}^2), \quad (10)$$

where

$$\begin{aligned}
-\frac{1}{2\sigma_{lk0}^2} &= -\frac{1}{2} \left(\frac{1}{1 - \hat{\rho}_e^2} \frac{\hat{\mathbf{R}}_{kk}^{(l)}}{\hat{\mathbf{s}}_{\gamma_{lk}}^2} + \frac{\beta_1^2}{\xi^2 \tau_1^2} + \frac{1}{\sigma_\gamma^2} \right), \\
\frac{\mu_{lk0}}{\sigma_{lk0}^2} &= \frac{1}{1 - \hat{\rho}_e^2} \left\{ \frac{\hat{\gamma}_{lk}}{\hat{\mathbf{s}}_{\gamma_{lk}}^2} - \sum_{j \neq k} \left(\frac{\hat{\mathbf{R}}_{jk}^{(l)} \gamma_{lj}}{\hat{\mathbf{s}}_{\gamma_{lj}}^2} \right) \frac{1}{\hat{\mathbf{s}}_{\gamma_{lk}}} - \hat{\rho}_e \frac{\hat{\Gamma}_{lk}}{\hat{\mathbf{s}}_{\gamma_{lk}} \hat{\mathbf{s}}_{\Gamma_{lk}}} + \left(\sum_{i=1}^p \frac{\hat{\mathbf{R}}_{ik}^{(l)} \Gamma_{li}}{\hat{\mathbf{s}}_{\Gamma_{li}}^2} \right) \frac{\hat{\rho}_e}{\hat{\mathbf{s}}_{\gamma_{lk}}} \right\} + \frac{\beta_1 \Gamma_{lk}}{\xi^2 \tau_1^2}, \\
-\frac{1}{2\sigma_{lk1}^2} &= -\frac{1}{2} \left(\frac{1}{1 - \hat{\rho}_e^2} \frac{\hat{\mathbf{R}}_{kk}^{(l)}}{\hat{\mathbf{s}}_{\gamma_{lk}}^2} + \frac{\beta_2^2}{\tau_2^2} + \frac{1}{\sigma_\gamma^2} \right), \\
\frac{\mu_{lk1}}{\sigma_{lk1}^2} &= \frac{1}{1 - \hat{\rho}_e^2} \left\{ \frac{\hat{\gamma}_{lk}}{\hat{\mathbf{s}}_{\gamma_{lk}}^2} - \sum_{j \neq k} \left(\frac{\hat{\mathbf{R}}_{jk}^{(l)} \gamma_{lj}}{\hat{\mathbf{s}}_{\gamma_{lj}}^2} \right) \frac{1}{\hat{\mathbf{s}}_{\gamma_{lk}}} - \hat{\rho}_e \frac{\hat{\Gamma}_{lk}}{\hat{\mathbf{s}}_{\gamma_{lk}} \hat{\mathbf{s}}_{\Gamma_{lk}}} + \left(\sum_{i=1}^p \frac{\hat{\mathbf{R}}_{ik}^{(l)} \Gamma_{li}}{\hat{\mathbf{s}}_{\Gamma_{li}}^2} \right) \frac{\hat{\rho}_e}{\hat{\mathbf{s}}_{\gamma_{lk}}} \right\} + \frac{\beta_2 \Gamma_{lk}}{\tau_2^2}.
\end{aligned}$$

Conditioning on the data and other parameters, both β_1 and β_2 have normal distributions,
i.e.,

$$\beta_1 | \mathbf{\Gamma}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \tau_1^2, \xi^2 \sim \mathcal{N}(\mu_{\beta_1}, \sigma_{\beta_1}^2), \quad \beta_2 | \mathbf{\Gamma}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \tau_2^2 \sim \mathcal{N}(\mu_{\beta_2}, \sigma_{\beta_2}^2), \quad (11)$$

where

$$\left\{ \begin{array}{l} -\frac{1}{2\sigma_{\beta_1}^2} = -\frac{1}{2} \frac{\sum_{l=1}^L (1-\eta_l) \boldsymbol{\gamma}_l^\top \boldsymbol{\gamma}_l}{\xi^2 \tau_1^2}, \\ \frac{\mu_{\beta_1}}{\sigma_{\beta_1}^2} = \frac{\sum_{l=1}^L (1-\eta_l) \boldsymbol{\Gamma}_l^\top \boldsymbol{\gamma}_l}{\xi^2 \tau_1^2}, \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{l} -\frac{1}{2\sigma_{\beta_2}^2} = -\frac{1}{2} \frac{\sum_{l=1}^L \eta_l \boldsymbol{\gamma}_l^\top \boldsymbol{\gamma}_l}{\tau_2^2}, \\ \frac{\mu_{\beta_2}}{\sigma_{\beta_2}^2} = \frac{\sum_{l=1}^L \eta_l \boldsymbol{\Gamma}_l^\top \boldsymbol{\gamma}_l}{\tau_2^2}. \end{array} \right.$$

The conditional posterior distributions of σ_γ^2 , τ_1^2 , τ_2^2 and ξ^2 are all inverse-gamma,i.e.,

$$\sigma_\gamma^2 | \boldsymbol{\gamma}, a_{\boldsymbol{\gamma}}, b_{\boldsymbol{\gamma}} \sim \mathcal{IG}(\tilde{a}_{\boldsymbol{\gamma}}, \tilde{b}_{\boldsymbol{\gamma}}), \quad \tau_1^2 | \mathbf{\Gamma}, \boldsymbol{\gamma}, \xi^2, a_{\tau_1}, b_{\tau_1} \sim \mathcal{IG}(\tilde{a}_{\tau_1}, \tilde{b}_{\tau_1}), \quad (12)$$

$$\tau_2^2 | \mathbf{\Gamma}, \boldsymbol{\gamma}, a_{\tau_2}, b_{\tau_2} \sim \mathcal{IG}(\tilde{a}_{\tau_2}, \tilde{b}_{\tau_2}), \quad \xi^2 | \mathbf{\Gamma}, \boldsymbol{\gamma}, \beta_1, \boldsymbol{\eta} \sim \mathcal{IG}(a_\xi, b_\xi), \quad (13)$$

$$\begin{aligned}
\text{where } \tilde{a}_{\boldsymbol{\gamma}} &= a_{\boldsymbol{\gamma}} + \frac{p}{2}, \quad \tilde{b}_{\boldsymbol{\gamma}} = b_{\boldsymbol{\gamma}} + \frac{\sum_{l=1}^L \boldsymbol{\gamma}_l^\top \boldsymbol{\gamma}_l}{2}, \quad \tilde{a}_{\tau_1} = a_{\tau_1} + \frac{\sum_{l=1}^L (1-\eta_l) p_l}{2}, \quad \tilde{b}_{\tau_1} = b_{\tau_1} + \frac{\sum_{l=1}^L (1-\eta_l) (\boldsymbol{\Gamma}_l - \beta_1 \boldsymbol{\gamma}_l)^\top (\boldsymbol{\Gamma}_l - \beta_1 \boldsymbol{\gamma}_l)}{2\xi^2}, \\
\tilde{a}_{\tau_2} &= a_{\tau_2} + \frac{\sum_{l=1}^L \eta_l p_l}{2}, \quad \tilde{b}_{\tau_2} = b_{\tau_2} + \frac{\sum_{l=1}^L \eta_l (\boldsymbol{\Gamma}_l - \beta_2 \boldsymbol{\gamma}_l)^\top (\boldsymbol{\Gamma}_l - \beta_2 \boldsymbol{\gamma}_l)}{2}, \quad a_\xi = \frac{1}{2} \sum_{l=1}^L (1 - \eta_l) p_l \text{ and} \\
b_\xi &= \frac{1}{2\tau_1^2} \sum_{l=1}^L (1 - \eta_l) (\boldsymbol{\Gamma}_l - \beta_1 \boldsymbol{\gamma}_l)^\top (\boldsymbol{\Gamma}_l - \beta_1 \boldsymbol{\gamma}_l).
\end{aligned}$$

The conditional posterior of ω is also a Beta distribution:

$$\omega|\boldsymbol{\eta}, a, b \sim \text{Beta} \left(a + \sum_{l=1}^L \eta_l, b + \sum_{l=1}^L (1 - \eta_l) \right). \quad (14)$$

The conditional probability of $\boldsymbol{\eta}$ given $\boldsymbol{\Gamma}$ can be found using Bayes' theorem

$$\begin{aligned} \Pr(\eta_l = 1 | \boldsymbol{\Gamma}_k) &= \frac{\Pr(\eta_l = 1) \Pr(\boldsymbol{\Gamma}_l | \eta_l = 1)}{\Pr(\eta_l = 0) \Pr(\boldsymbol{\Gamma}_l | \eta_l = 0) + \Pr(\eta_l = 1) \Pr(\boldsymbol{\Gamma}_l | \eta_l = 1)} \\ &= \frac{\omega \prod_{k=1}^{p_l} \mathcal{N}(\beta_2 \gamma_{kl}, \tau_2^2)}{\omega \prod_{k=1}^{p_l} \mathcal{N}(\beta_2 \gamma_{kl}, \tau_2^2) + (1 - \omega) \prod_{k=1}^{p_l} \mathcal{N}(\beta_1 \gamma_{kl}, \xi^2 \tau_1^2)}. \end{aligned} \quad (15)$$

Algorithm 2 The parallel Gibbs sampler algorithm for correlated SNPs

```

1: Initialize:

2: repeat
3:   for  $l = 1$  to  $L$  do
4:     for  $k = 1$  to  $p_l$  do
5:       Sample  $\Gamma_{lk}$  using (9).
6:     end for
7:     for  $k = 1$  to  $p_l$  do
8:       Sample  $\gamma_{lk}$  using (10).
9:     end for
10:    Sample  $\eta_l$  using (15).
11:  end for
12:  Sample  $\beta_1$  and  $\beta_2$  using (11).
13:  Sample  $\sigma_\gamma^2$ ,  $\tau_1^2$ ,  $\tau_2^2$  and  $\xi^2$  using (12).
14:  Sample  $\omega$  using (14).
15: until Maximum iteration reached

```

Working Parallel

Supplementary Note 2

Pseudo-code

In this section, we present the pseudo-code for the algorithms using correlated SNPs.

Comment lines are preceded by % %.

$\mathbf{a}[i]$ denotes the i -th element of a vector \mathbf{a} . $\mathbf{a}[s : e]$ denotes a subvector of \mathbf{a} , spanning from the s -th to the e -th element. $\mathbf{A}[:, i]$ denotes the i -th column of a matrix \mathbf{A} . $\mathbf{a} \circ \mathbf{b}$ denotes element-wise multiplication. \mathbf{a}/\mathbf{b} denotes element-wise division. In addition, we denote the

diagonal of a matrix \mathbf{A} by $\text{diag}(\mathbf{A})$. A square diagonal matrix with the elements of vector \mathbf{a} on the main diagonal is denoted by $\text{diagmat}(\mathbf{a})$.

% % Input observations and prior parameters.

Input: $\hat{\gamma}, \hat{\Gamma}, \hat{s}_{\gamma}^2, \hat{s}_{\Gamma}^2, \hat{\rho}_e, \hat{\mathbf{R}}, \mathbf{G}; a_{\gamma} \leftarrow 0, b_{\gamma} \leftarrow 0, a_{\tau 1} \leftarrow 0, b_{\tau 1} \leftarrow 0, a_{\tau 2} \leftarrow 0, b_{\tau 2} \leftarrow 0, a \leftarrow 2, b \leftarrow L$.

% % \mathbf{G} is a $L \times 2$ matrix, which stores the group information, start with the first element and end with the second element of each row.

Output: β_1, β_2, η .

% % Initialize latent variables.

$\omega \leftarrow 0.1, \beta_1 \leftarrow 0.01, \beta_2 \leftarrow 0.01, \sigma_{\gamma}^2 \leftarrow 0.01, \tau_1^2 \leftarrow 0.01, \tau_2^2 \leftarrow 0.01, \xi^2 \leftarrow 0.01, \gamma = 0.01 * \mathbf{1}_P, \Gamma = \mathbf{0.01} * \mathbf{1}_P, \eta = \mathbf{0}_L$.

$\gamma_{\eta^0} \leftarrow \mathbf{0}_L; \gamma_{\eta^1} \leftarrow \mathbf{0}_L; \gamma_{\Gamma \eta^0} \leftarrow \mathbf{0}_L; \gamma_{\Gamma \eta^1} \leftarrow \mathbf{0}_L; \eta_0^L \leftarrow \mathbf{0}_L \eta_1^L \leftarrow \mathbf{0}_L; t_{\tau_1^2} \leftarrow \mathbf{0}_L; t_{\tau_2^2} \leftarrow \mathbf{0}_L;$

$\rho_1^* \leftarrow 1/(1 - \hat{\rho}_e * \hat{\rho}_e); \rho_2^* \leftarrow \hat{\rho}_e/(1 - \hat{\rho}_e * \hat{\rho}_e); \rho_{\gamma \Gamma} \leftarrow \rho_1^* \hat{s}_{\gamma l} / \hat{s}_{\Gamma l}; \rho_{\Gamma \gamma} \leftarrow \rho_1^* \hat{s}_{\Gamma l} / \hat{s}_{\gamma l};$

$\hat{\zeta}_{\gamma} \leftarrow \rho_2^* \cdot \hat{\gamma}_l / (\hat{s}_{\Gamma l} \circ \hat{s}_{\gamma l}); \hat{\zeta}_{\Gamma} \leftarrow \rho_2^* \cdot \hat{\Gamma}_l / (\hat{s}_{\Gamma l} \circ \hat{s}_{\gamma l});$

for $i = 1$ to number of MCMC iterations **do**

$logw \leftarrow \log(\omega/(1 - \omega))$

for $l = 1 \in 1, \dots, L$ **do**

% % Parameters for updating Γ_l .

$s_l \leftarrow \mathbf{G}[l, 1]; e_l \leftarrow \mathbf{G}[l, 2]; \hat{\mathbf{R}}^{(l)} \leftarrow \hat{\mathbf{R}}[s_l : e_l, s_l : e_l]; \mathbf{d}_l = \text{diag}(\hat{\mathbf{R}}^{(l)});$

$\hat{s}_{\gamma l}^2 \leftarrow \hat{s}_{\gamma}^2[s_l : e_l]; \hat{s}_{\Gamma l}^2 \leftarrow \hat{s}_{\Gamma}^2[s_l : e_l]; \hat{\gamma}_l \leftarrow \hat{\gamma}[s_l : e_l]; \hat{\Gamma}_l \leftarrow \hat{\Gamma}[s_l : e_l];$

$\hat{\zeta}_{\gamma l} \leftarrow \hat{\zeta}_{\gamma}[s_l : e_l]; \hat{\zeta}_{\Gamma l} \leftarrow \hat{\zeta}_{\Gamma}[s_l : e_l];$

$\rho_{\gamma \Gamma}^l \leftarrow \rho_{\gamma \Gamma}[s_l : e_l]; \rho_{\gamma \Gamma}^l \leftarrow \rho_{\Gamma \gamma}[s_l : e_l];$

$\gamma_l \leftarrow \gamma[s_l : e_l]; \Gamma_l \leftarrow \Gamma[s_l : e_l]; p_l \leftarrow \text{length}(\gamma_l);$

$\delta_{\Gamma l} \leftarrow \text{diag}(\rho_1 \hat{\mathbf{R}}^{(l)} * \text{diagmat}(1/\hat{s}_{\Gamma l})); \delta_{\gamma l} \leftarrow \text{diag}(\rho_1 \hat{\mathbf{R}}^{(l)} * \text{diagmat}(1/\hat{s}_{\gamma l}));$

$\Delta_{\Gamma l} \leftarrow \rho_1^* \hat{\mathbf{R}}^{(l)} * \text{diagmat}(1/\hat{s}_{\Gamma l}); \Delta_{\gamma l} \leftarrow \rho_1^* \hat{\mathbf{R}}^{(l)} * \text{diagmat}(1/\hat{s}_{\gamma l});$

```

 $v_{\Gamma 1}^2 \leftarrow 1/(\rho_2^* \widehat{\boldsymbol{\Gamma}}_l / (\widehat{\mathbf{s}}_{\gamma l} \circ \widehat{\mathbf{s}}_{\Gamma l}) + 1/(\tau_1^2 \xi^2)); v_{\Gamma 2}^2 \leftarrow 1/(\rho_2^* \widehat{\boldsymbol{\Gamma}}_l / (\widehat{\mathbf{s}}_{\gamma l} \circ \widehat{\mathbf{s}}_{\Gamma l}) + 1/\tau_2^2);$ 
 $v_{\gamma 1}^2 \leftarrow 1/(\rho_2^* \widehat{\boldsymbol{\gamma}}_l / (\widehat{\mathbf{s}}_{\gamma l} \circ \widehat{\mathbf{s}}_{\Gamma l}) + \beta_1 * \beta_1 / (\tau_1^2 \xi^2) + 1/\sigma_\gamma^2); v_{\gamma 2}^2 \leftarrow 1/(\rho_2^* \widehat{\boldsymbol{\gamma}}_l / (\widehat{\mathbf{s}}_{\gamma l} \circ \widehat{\mathbf{s}}_{\Gamma l}) + \beta_2 *$ 
 $\beta_2 / \tau_2^2 + 1/\sigma_\gamma^2);$ 
 $\mu_{\Gamma l} \leftarrow \rho_1^* \widehat{\mathbf{R}}^{(l)} * \text{diagmat}(1/\mathbf{s}_{\Gamma l}) * \boldsymbol{\Gamma}_l; \mu_{\gamma l} \leftarrow \rho_1^* \widehat{\mathbf{R}}^{(l)} * \text{diagmat}(1/\mathbf{s}_{\gamma l}) * \boldsymbol{\gamma}_l;$ 
% % Sample  $\boldsymbol{\Gamma}_l$  in group  $l$ .
if  $\eta[l] = 1$  then
    for  $k = 1 \in 1, \dots, p_l$  do
         $t_1 \leftarrow \mu_{\Gamma l} - \boldsymbol{\Gamma}_l[k] \cdot \Delta_{\Gamma l}[k];$ 
         $\mu_{\Gamma l}^k \leftarrow \mu_{\Gamma l}[k] - \delta_{\Gamma l}[k] \cdot \boldsymbol{\Gamma}_l[k];$ 
         $\tilde{m}_{l1} \leftarrow (\rho_{\Gamma \gamma}^l[k] - \mu_{\Gamma l}^k / \widehat{\boldsymbol{\Gamma}}_l[k] - \widehat{\boldsymbol{\zeta}}_{\gamma l}[k] + \rho \cdot \mu_{\gamma l}[k] / \widehat{\boldsymbol{\Gamma}}_l[k] + \beta_2 \cdot \Gamma_l[k] / \tau_2^2)) / v_{\Gamma 2}^2[k];$ 
        % % Draw  $\boldsymbol{\Gamma}_l[k]$  from normal distribution.
         $\boldsymbol{\Gamma}_l[k] \sim \mathcal{N}(\tilde{m}_{l1}, v_{\Gamma 1}^2[k]);$ 
         $\mu_{\Gamma l} \leftarrow t_1 + \boldsymbol{\Gamma}_l[k] \cdot \Delta_{\Gamma l}[k];$ 
    end for
    else
        for  $k = 1 \in 1, \dots, p_l$  do
             $t_0 \leftarrow \mu_{\Gamma l} - \boldsymbol{\Gamma}_l[k] \cdot \Delta_{\Gamma l}[k];$ 
             $\mu_{\Gamma l}^k \leftarrow \mu_{\Gamma l}[k] - \delta_{\Gamma l}[k] \cdot \boldsymbol{\Gamma}_l[k];$ 
             $\tilde{m}_{l0} \leftarrow (\rho_{\Gamma \gamma}^l[k] - \mu_{\Gamma l}^k / \widehat{\boldsymbol{\Gamma}}_l[k] - \widehat{\boldsymbol{\zeta}}_{\gamma l}[k] + \rho \cdot \mu_{\gamma l}[k] / \widehat{\boldsymbol{\Gamma}}_l[k] + \beta_1 \cdot \Gamma_l[k] / (\tau_1^2 \xi^2)) / v_{\Gamma 1}^2[k];$ 
            % % Draw  $\boldsymbol{\Gamma}_l[k]$  from normal distribution.
             $\boldsymbol{\Gamma}_l[k] \sim \mathcal{N}(\tilde{m}_{l0}, v_{\Gamma 0}^2[k]);$ 
             $\mu_{\Gamma l} \leftarrow t_0 + \boldsymbol{\Gamma}_l[k] \cdot \Delta_{\Gamma l}[k];$ 
        end for
    end if
% % Sample  $\boldsymbol{\gamma}_l$  in group  $l$ .

```

```

if  $\eta[l] = 1$  then

    for  $k = 1 \in 1, \dots, p_l$  do

         $t_1 \leftarrow \mu_{\gamma l} - \gamma_l[k] \cdot \Delta_{\gamma l}[k];$ 
         $\mu_{\gamma l}^k \leftarrow \mu_{\gamma l}[k] - \delta_{\gamma l}[k] \cdot \gamma_l[k];$ 
         $m_{l1} \leftarrow (\rho_{\gamma \Gamma}^l[k] - \mu_{\gamma l}^k / \hat{\gamma}_l[k] - \zeta_{\Gamma l}[k] + \rho \cdot \mu_{\Gamma l}[k] / \hat{\gamma}_l[k] + \beta_2 \cdot \gamma_l[k] / \tau_2^2) / v_{\gamma 2}^2[k];$ 
        % % Draw  $\gamma_l[k]$  from normal distribution.
         $\gamma_l[k] \sim \mathcal{N}(m_{l1}, v_{\gamma 2}^2[k]);$ 
         $\mu_{\gamma l} \leftarrow t_1 + \gamma_l[k] \cdot \Delta_{\gamma l}[k];$ 

    end for

    else

        for  $k = 1 \in 1, \dots, p_l$  do

             $t_0 \leftarrow \mu_{\gamma l} - \gamma_l[k] \cdot \Delta_{\gamma l}[k];$ 
             $\mu_{\gamma l}^k \leftarrow \mu_{\gamma l}[k] - \delta_{\gamma l}[k] \cdot \gamma_l[k];$ 
             $m_{l0} \leftarrow (\rho_{\gamma \Gamma}^l[k] - \mu_{\gamma l}^k / \hat{\gamma}_l[k] - \zeta_{\Gamma l}[k] + \rho \cdot \mu_{\Gamma l}[k] / \hat{\gamma}_l[k] + \beta_1 \cdot \gamma_l[k] / (\tau_1^2 \xi^2)) / v_{\gamma 1}^2[k];$ 
            % % Draw  $\gamma_l[k]$  from normal distribution.
             $\gamma_l[k] \sim \mathcal{N}(m_{l0}, v_{\gamma 1}^2[k]);$ 
             $\mu_{\gamma l} \leftarrow t_0 + \gamma_l[k] \cdot \Delta_{\gamma l}[k];$ 

        end for

    end if

    % Update  $\eta[l]$ .
     $\epsilon_2 \leftarrow \Gamma_l - \beta_2 \gamma_l; \quad \epsilon_1 \leftarrow \Gamma_l - \beta_1 \gamma_l;$ 
     $t_2 \leftarrow \log w - \text{sum}(\epsilon_2 \circ \epsilon_2) / (2\tau_2^2) - p_l \cdot \log(\tau_2^2) / 2 + \text{sum}(\epsilon_1 \circ \epsilon_1) / (2\tau_1^2 \xi^2) + p_1 \cdot \log(\tau_1^2 \xi^2) / 2.$ 
     $prob \leftarrow 1 / (1 + \exp(-t_2));$ 
     $\eta[l] \sim \text{Binom}(prob);$ 
    % Renew some terms for the following procedure.

```

```

 $\gamma_{\eta^0}[s_l : e_l] \leftarrow (1 - \eta[l]) \cdot sum(\gamma_l \circ \gamma_l);$ 
 $\gamma_{\eta^1}[s_l : e_l] \leftarrow \eta[l] \cdot sum(\gamma_l \circ \gamma_l);$ 
 $\gamma_{\Gamma_{\eta^0}}[s_l : e_l] \leftarrow (1 - \eta[l]) \cdot sum(\Gamma_l \circ \gamma_l);$ 
 $\gamma_{\Gamma_{\eta^1}}[s_l : e_l] \leftarrow \eta[l] \cdot sum(\Gamma_l \circ \gamma_l);$ 
 $\eta_0^L[s_l : e_l] \leftarrow p_l \cdot (1 - \eta[l]);$ 
 $\eta_1^L[s_l : e_l] \leftarrow p_l \cdot \eta[l];$ 
 $t_{\tau_1^2}[s_l : e_l] \leftarrow (1 - \eta[l]) \cdot sum((\Gamma_l - \beta_1 \cdot \gamma_l) \circ (\Gamma_l - \beta_1 \cdot \gamma_l));$ 
 $t_{\tau_2^2}[s_l : e_l] \leftarrow \eta[l] \cdot sum((\Gamma_l - \beta_2 \cdot \gamma_l) \circ (\Gamma_l - \beta_2 \cdot \gamma_l));$ 

end for

% % Draw  $\beta_1$  from normal distribution.

if sum( $\eta=p$ ) then
     $\beta_1 \leftarrow 0.$ 
else
     $\sigma_{\beta 1}^2 \leftarrow \tau_1^2 \xi^2 / sum(\gamma_{\eta^0})$ 
     $m_{\beta 1} \leftarrow \sigma_{\beta 1}^2 sum(\gamma_{\Gamma_{\eta^0}}) / (\tau_1^2 \xi^2)$ 
     $\beta_1 \sim \mathcal{N}(m_{\beta 1}, \sigma_{\beta 1}^2).$ 
end if

% % Draw  $\beta_2$  from normal distribution.

if sum( $\eta=0$ ) then
     $\beta_2 \leftarrow 0;$ 
else
     $\sigma_{\beta 2}^2 \leftarrow \tau_2^2 / sum(\gamma_{\eta^1})$ 
     $m_{\beta 2} \leftarrow \sigma_{\beta 2}^2 sum(\gamma_{\Gamma_{\eta^1}}) / \tau_2^2$ 
     $\beta_2 \sim \mathcal{N}(m_{\beta 2}, \sigma_{\beta 2}^2).$ 
end if

```

```

 $\tilde{a}_\gamma \leftarrow a_\gamma + p/2; \tilde{b}_\gamma \leftarrow b_\gamma + sum(\gamma \circ \gamma)/2.$ 
 $\tilde{a}_{\tau 1} \leftarrow a_{\tau 1} + sum(\eta_0^L)/2; \tilde{b}_{\tau 1} \leftarrow b_{\tau 1} + sum(t_{\tau_1^2})/(2\xi^2).$ 
 $\tilde{a}_{\tau 2} \leftarrow a_{\tau 2} + sum(\eta_1^L)/2; \tilde{b}_{\tau 2} \leftarrow b_{\tau 2} + sum(t_{\tau_2^2})/(2).$ 
 $a_\xi \leftarrow sum(\eta_0^L)/2; b_\xi \leftarrow sum(t_{\tau_1^2})/(2\tau_1^2).$ 
% % Draw  $\sigma_\gamma^2$  from Inverse-gamma distribution.
 $\sigma_\gamma^2 \sim \mathcal{IG}(\tilde{a}_\gamma, \tilde{b}_\gamma).$ 
% % Draw  $\tau_1^2$  from Inverse-gamma distribution.
 $\tau_1^2 \sim \mathcal{IG}(\tilde{a}_{\tau 1}, \tilde{b}_{\tau 1}).$ 
% % Draw  $\tau_2^2$  from Inverse-gamma distribution.
 $\tau_2^2 \sim \mathcal{IG}(\tilde{a}_{\tau 2}, \tilde{b}_{\tau 2}).$ 
% % Draw  $\xi^2$  from Inverse-gamma distribution.
 $\xi^2 \sim \mathcal{IG}(a_\xi, b_\xi).$ 
% % Draw  $\omega$  from Beta distribution.
 $\omega \sim \text{Beta}(a + sum(\eta), b + sum(1 - \eta)).$ 
% % Store  $\eta$  of each iteration into a matrix  $\mathbf{H}$ .
 $\mathbf{H}.col(i) \leftarrow \eta$ 
end for

```

Supplementary Note 3

Additional simulation results

3.1 The identifiability of causal effect

In this subsection, we discuss and evaluate the identifiability of causal effect. Following existing literature [1, 2], our model also assumes that all IVs could have potential uncorrelated pleiotropic effect while only a sparse proportion of IVs have correlated pleiotropic effect. As a consequence of the assumption, the variability of Γ_k is larger for the β_2 group of SNPs

than the β_1 group because of the existence of $\tilde{\alpha}_k$.

In Equation (7) of the main text, we have

$$\Gamma_k | \beta_1, \beta_2, \gamma_k, \eta_k, \tau_1^2, \tau_2^2 \sim \begin{cases} \mathcal{N}(\beta_1 \gamma_k, \tau_1^2), & \eta_k = 0 \\ \mathcal{N}(\beta_2 \gamma_k, \tau_2^2), & \eta_k = 1 \end{cases}, \quad (16)$$

where $\tau_1^2 = \sigma_\theta^2$ for IVs with potential UHP only and $\tau_2^2 = \sigma_\theta^2 + \sigma_\alpha^2$ with both potential UHP and CHP. In this formulation, $\tau_2^2 > \tau_1^2$. Since both τ_1^2 and τ_2^2 are model parameters, we can obtain their estimates using MCMC.

To evaluate the validity of the condition in different settings, we conducted the following simulations. For each block of IV sets l , we reparameterize γ_{lk} and α_{lk} as follows

$$\begin{aligned} \gamma_{lk} &\sim \mathcal{N}(0, \sigma_\gamma^2) \\ \alpha_{lk} &= \delta \cdot \gamma_{lk} + \tilde{\alpha}_{lk} \end{aligned}$$

where $\tilde{\alpha}_{lk} \sim \mathcal{N}(0, \sigma_\alpha^2)$ and $\gamma_{lk} \perp\!\!\!\perp \tilde{\alpha}_{lk}$. Thus, for IVs with no CHP effect, the causal effect from exposure to outcome is given by β_1 . For IVs with both UHP and CHP effects, the causal effect is $\beta_2 = \delta + \beta_1$, and δ is the difference/bias in estimation due to CHP. To assess the identifiability of causal effects, we vary the value of δ and conduct the simulations with 100 replications. We examine the proportion of sets of estimates with $\hat{\tau}_1^2 < \hat{\tau}_2^2$. We set the true causal effect equals to 1. In Table 1, we show that when the sparsity levels of IVs with CHP is sparse (0.1) or moderately sparse (0.3), using the condition $\hat{\tau}_1^2 < \hat{\tau}_2^2$ to identify the causal effect $\hat{\beta}_1$ works well. However, when the proportion of IVs with CHP becomes non-sparse (0.5) and when the difference between β_1 and β_2 (i.e., δ) is large (1 or -1), the identification has some issues and the estimated β_1 start to show biases. It should be noted that when the proportion of IVs with CHP is non-sparse, CHP and UHP effects becomes statistically unidentifiable. Thus, in existing literature, it is commonly assumed that the

δ	Proportions of IV with CHP: 0.1			Proportions of IV with CHP: 0.3			Proportions of IV with CHP: 0.5		
	$\hat{\beta}_1$	$\hat{\beta}_2$	Prop. of replications with $\hat{\tau}_1^2 < \hat{\tau}_2^2$ (%)	$\hat{\beta}_1$	$\hat{\beta}_2$	Prop. of replications with $\hat{\tau}_1^2 < \hat{\tau}_2^2$ (%)	$\hat{\beta}_1$	$\hat{\beta}_2$	Prop. of replications with $\hat{\tau}_1^2 < \hat{\tau}_2^2$ (%)
0.0010	1.004	1.007	100	0.996	1.001	100	1.006	0.491	71
0.1000	0.997	1.111	100	1.003	1.083	99	1.020	0.513	73
1.0000	0.998	1.726	100	1.046	1.498	100	1.148	0.657	55
-1.0000	0.992	0.299	100	0.952	0.482	100	0.872	0.161	72

Supplementary Table 1: An evaluation of the identifiability of causal effects in various simulation settings, using the condition $\hat{\tau}_1^2 < \hat{\tau}_2^2$. Here $\sigma_\alpha^2 = 0.1$, $h_\gamma^2 = 0.1$, $h_\alpha^2 = 0.05$, $h_\theta^2 = 0.05$, $r = 0.4$, $p = 1,000$.

proportion of IVs with CHP is sparse or moderately sparse [3, 4, 1].

As a conclusion, when the proportion of IVs with CHP effect is sparse or moderately sparse (< 0.5), the condition of $\hat{\tau}_1^2 < \hat{\tau}_2^2$ works very well in identifying the true causal effects. When the proportion is non-sparse (0.5) and the confounding bias (δ) is large, the identification condition may lead to biased estimation.

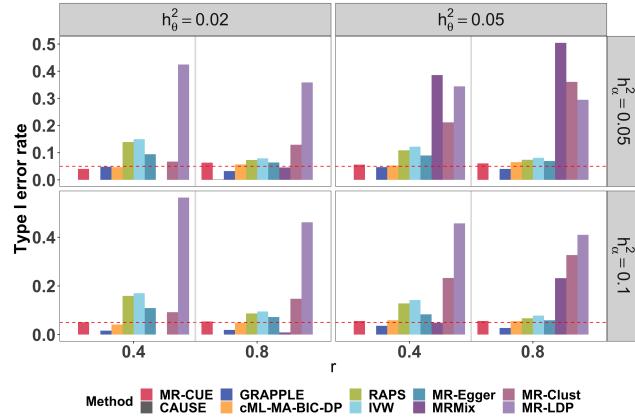
3.2 Simulation results for different p and $\rho_{\alpha\gamma}$

In this subsection, we provide additional results on type I error rate with different number of IVs ($p = 2,000$) and a stronger correlation between IV-to-exposure and CHP effects ($\rho_{\alpha\gamma} = 0.8$), see Fig. 1. To compare the bias of causal effect estimate derived from each method, we show the box plot of point estimates for each method in Fig. 2, with the true causal effect $\beta_1 = 1$. We further summarize the mean bias of causal effect estimates from MR-CUE and other methods in Table 2. The mean bias of MR-CUE is the smallest among all methods. Also, We compare the power of each method by varying $h_\gamma^2 \in \{0, 10^{-3.5}, 10^{-3}, 10^{-2.5}, 10^{-2}, 10^{-1.5}, 0.1\}$ with different setting, i.e., $h_\theta^2 \in \{0.02, 0.05\}$, $h_\alpha \in \{0.05, 0.1\}$, $r \in \{0.4, 0.8\}$ and $p \in \{1000, 2000\}$. One can found the corresponding results in Fig. 3-6. In addition, we report the false discovery rate (FDR) with $\beta_1 = 0$ and $\beta_1 = 0.1$, respectively. The corresponding result can be found in Fig. 7. We evaluate the performance of identification of IV with CHP ($\eta = 1$) using the area under the receiver operating characteristic (ROC) curve (AUC), see

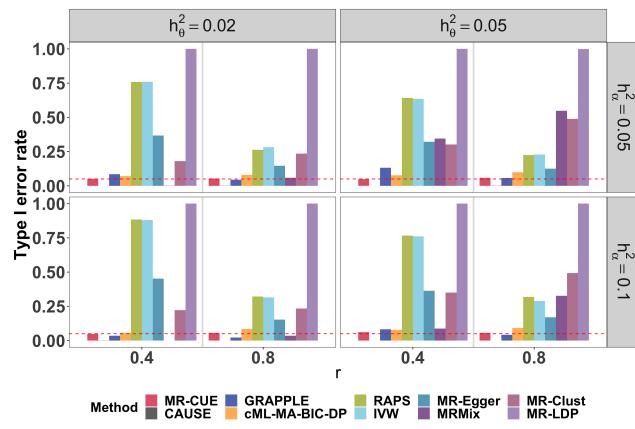
Fig. 8.

Method	$h_\theta^2 = 0.02, h_\alpha^2 = 0.05$	$h_\theta^2 = 0.05, h_\alpha^2 = 0.1$
MR-CUE	0.071	-0.149
GRAPPLE	-0.207	-0.780
MRMix	1.940	5.470
cML-MA-BIC-DP	2.667	7.035
RAPS	4.446	6.243
MR-Clust	-5.251	-7.092
MR-LDP	-9.620	-30.535
IVW	-11.301	-8.271
MR-Egger	-11.708	-9.026
CAUSE	-24.063	-22.107

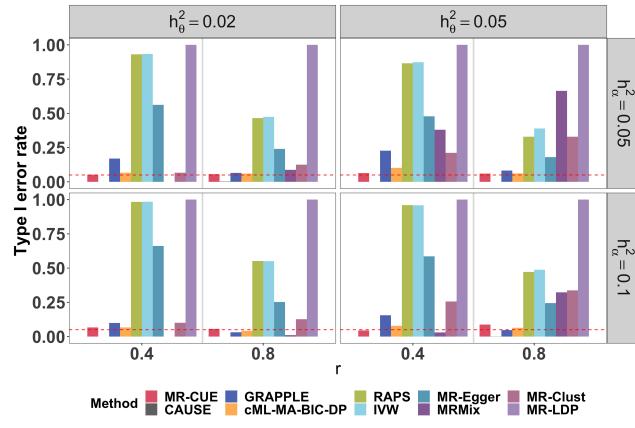
Supplementary Table 2: Mean biases of causal effect estimates over 100 replications for competing methods, with $h_\gamma^2 = 0.1, r = 0.4, \rho_{\alpha\gamma} = 0.2, p = 2,000$, all in %.



(a) $p = 2,000, \rho_{\alpha\gamma} = 0.2$

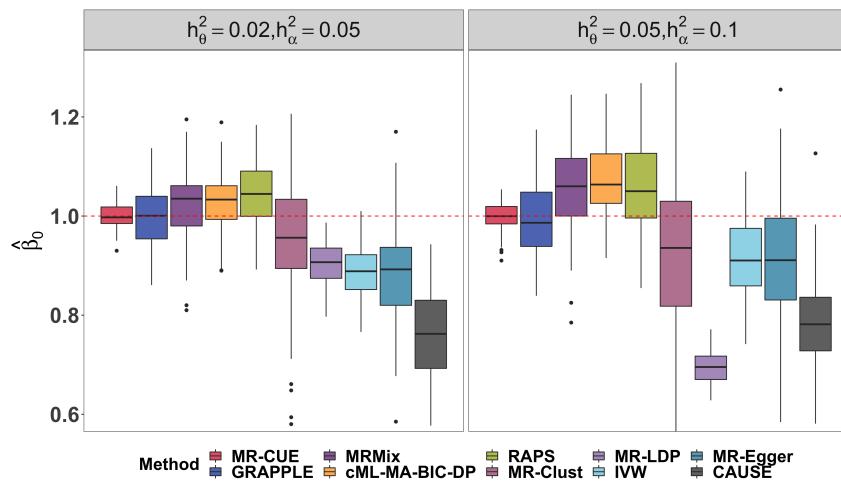


(b) $p = 1,000, \rho_{\alpha\gamma} = 0.8$

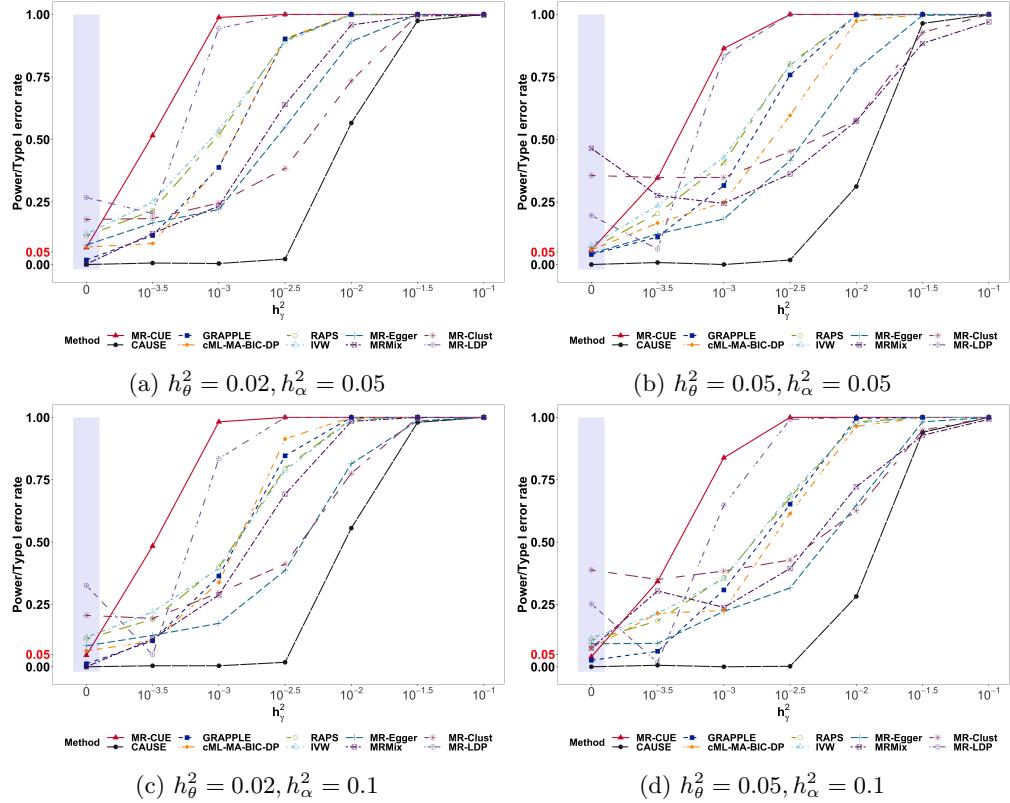


(c) $p = 2,000, \rho_{\alpha\gamma} = 0.8$

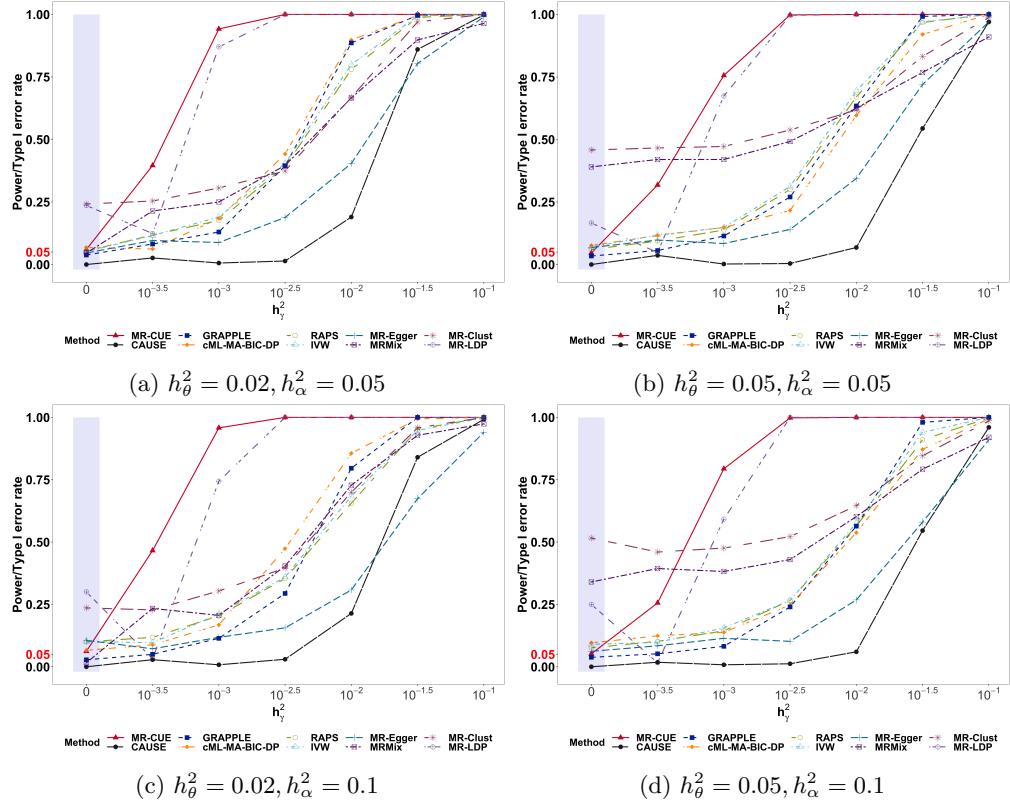
Supplementary Figure 1: Evaluation of type I error rate with $\beta_1 = 0$. The number of replicates is 1,000. Note that we evaluate the type I error rate for CAUSE, GRAPPLE, cML-MA-BIC-DP, RAPS, IVW, MR-Egger, MRMix and MR-Clust using independent SNPs.



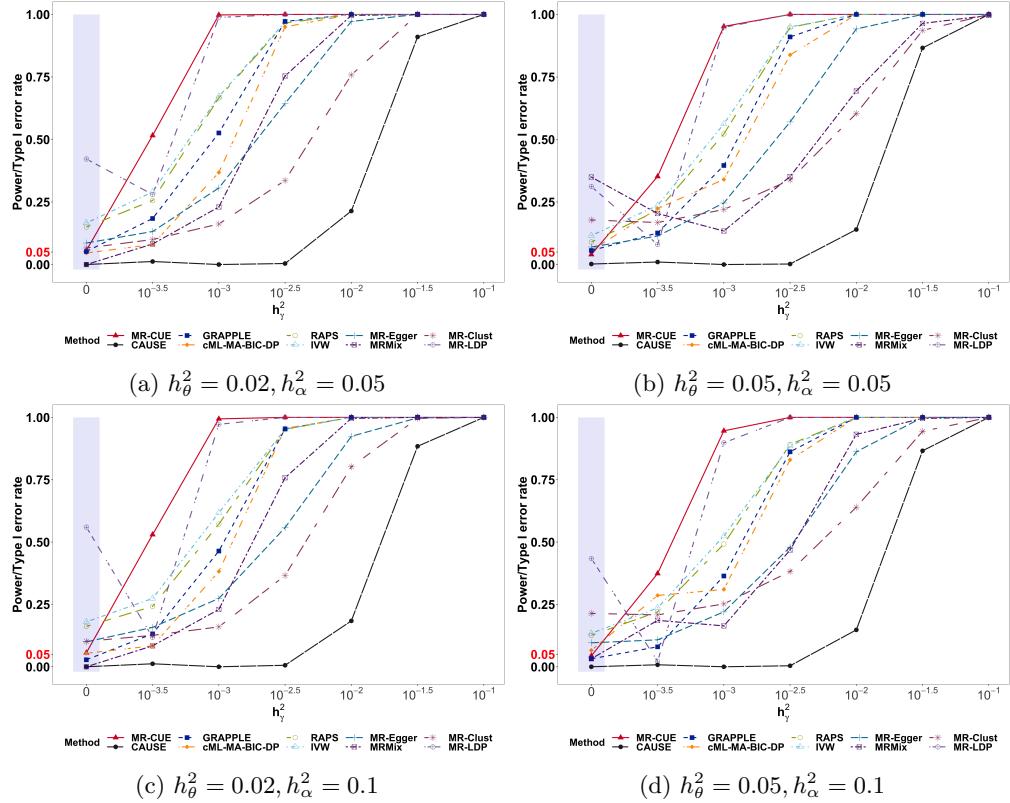
Supplementary Figure 2: The boxplots of point estimates over 100 replications for competing methods, with $h_\gamma^2 = 0.1$, $r = 0.4$, $\rho_{\alpha\gamma} = 0.2$, $p = 2,000$ and sample sizes $n_x = 50,000$, $n_y = 50,000$. The bounds of the boxes represent 25% and 75% quantiles and the center lines the median values. The top and bottom whiskers are set to the highest/lowest value of the dataset that are included in the $1.5 \times$ the interquartile range, outliers are defined as data points that are located outside the whiskers.



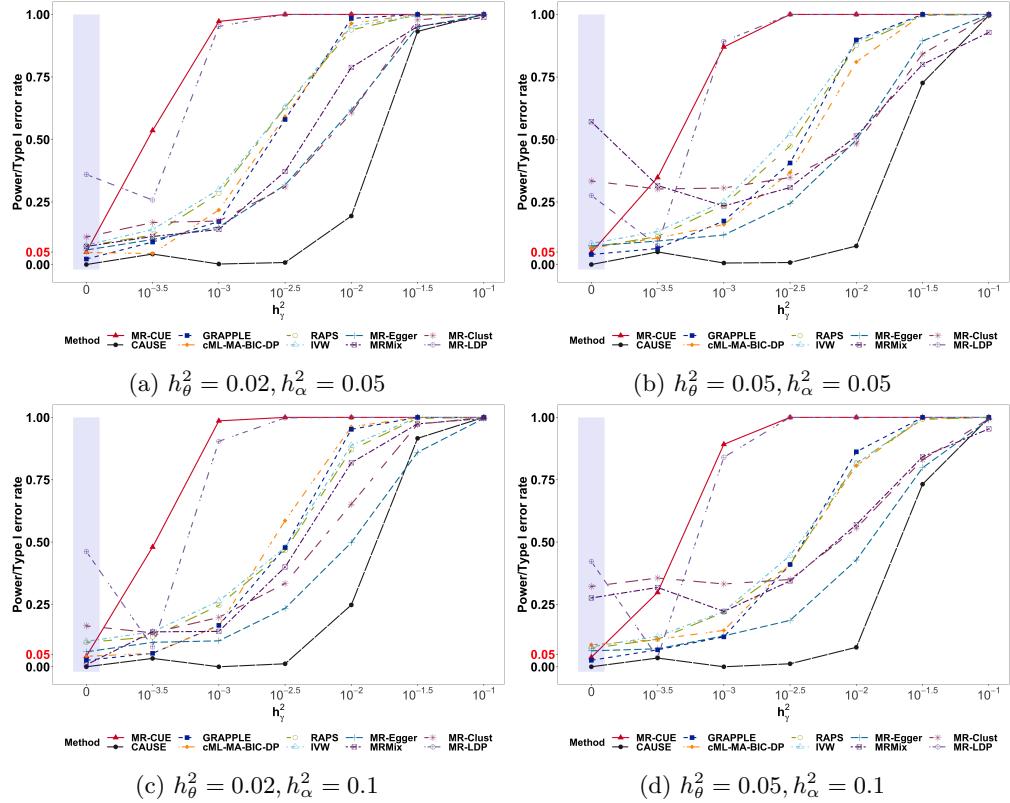
Supplementary Figure 3: Power comparison with $\rho_{\alpha\gamma} = 0.2, r = 0.4$ and $p = 1,000$. The number of replicates is 500. Note that we evaluate the type I error rate for CAUSE, GRAPPLE, cML-MA-BIC-DP, RAPS, IVW, MR-Egger, MRMix and MR-Clust using independent SNPs.



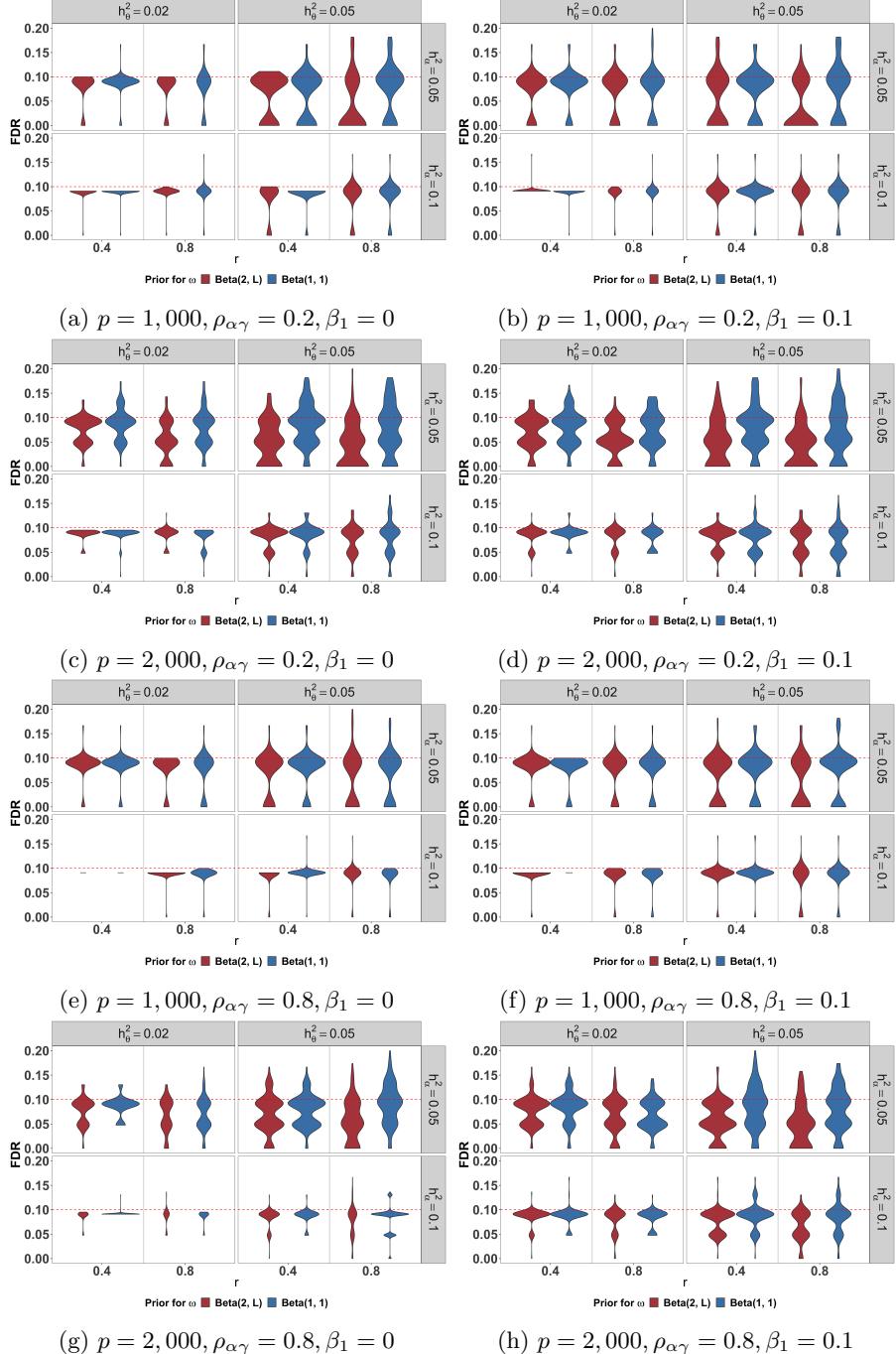
Supplementary Figure 4: Power comparison with $\rho_{\alpha\gamma} = 0.2, r = 0.8$ and $p = 1,000$. The number of replicates is 500. Note that we evaluate the type I error rate for CAUSE, GRAPPLE, cML-MA-BIC-DP, RAPS, IVW, MR-Egger, MRMix and MR-Clust using independent SNPs.



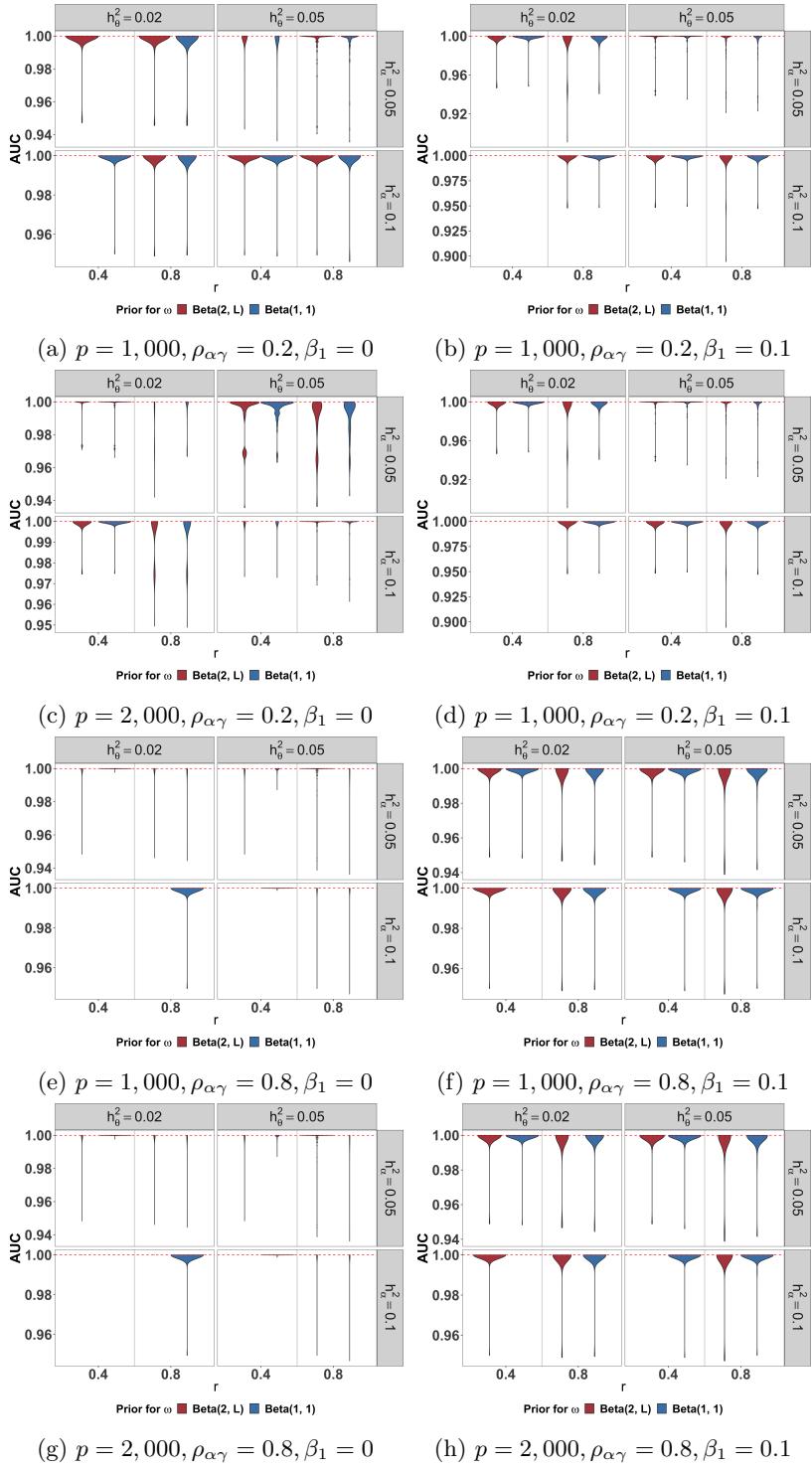
Supplementary Figure 5: Power comparison with $\rho_{\alpha\gamma} = 0.2, r = 0.4$ and $p = 2,000$. The number of replicates is 500. Note that we evaluate the type I error rate for CAUSE, GRAPPLE, cML-MA-BIC-DP, RAPS, IVW, MR-Egger, MRMix and MR-Clust using independent SNPs.



Supplementary Figure 6: Power comparison with $\rho_{\alpha\gamma} = 0.2, r = 0.8$ and $p = 2,000$. The number of replicates is 500. Note that we evaluate the type I error rate for CAUSE, GRAPPLE, cML-MA-BIC-DP, RAPS, IVW, MR-Egger, MRMix and MR-Clust using independent SNPs.



Supplementary Figure 7: Evaluation of FDR with $\beta_1 = 0$ (left panel) and $\beta_1 = 0.1$ (right panel). The number of replicates is 100.



Supplementary Figure 8: Evaluation of AUC using different prior for ω with $\beta_1 = 0$ (left panel) and $\beta_1 = 0.1$ (right panel). The number of replicates is 100.

3.3 Generation of summary statistics for multiple confounding pathways

In this subsection, we introduce the detailed simulation settings related to the scenario of multiple confounding pathways. Specifically, we simulate 5 groups of IVs being associated with 5 different confounders with varying effect sizes. In details, for the l -th block, we simulate

$$\gamma_{lk} \sim \mathcal{N}(0, \sigma_\gamma^2)$$

$$\alpha_{lk} = \delta \cdot \gamma_{lk} + \tilde{\alpha}_{lk}$$

where $\tilde{\alpha}_{lk} \sim \mathcal{N}(0, \sigma_\alpha^2)$ and $\gamma_{lk} \perp\!\!\!\perp \tilde{\alpha}_{lk}$. The five groups of IVs are associated with those five confounders with effects, $(\sigma_{\alpha 1}^2 = 0.02, \delta_1 = 0.02)$, $(\sigma_{\alpha 2}^2 = 0.01, \delta_2 = 0.05)$, $(\sigma_{\alpha 3}^2 = 0.01, \delta_3 = 0.1)$, $(\sigma_{\alpha 4}^2 = 0.005, \delta_4 = 0.5)$, and $(\sigma_{\alpha 5}^2 = 0.005, \delta_5 = 0.8)$. We consider α_k to be sparse, i.e., only 10% of α_k 's are sampled from the bivariate normal distribution and the others were zero. For UHP, we assume θ_k to be dense and follow an independent normal distribution, $\mathcal{N}(0, \sigma_\theta^2)$.

Once we have $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$, we adopte similar procedure to generate the individual-level data,

$$\begin{aligned}\mathbf{x}_x &= \mathbf{G}_x \boldsymbol{\gamma} + \mathbf{U}_x \boldsymbol{\psi}_x + \boldsymbol{\epsilon}_{x_x}, \\ \mathbf{x}_y &= \mathbf{G}_y \boldsymbol{\gamma} + \mathbf{U}_y \boldsymbol{\psi}_x + \boldsymbol{\epsilon}_{x_y}, \\ \mathbf{y} &= \beta_1 \mathbf{x}_y + \mathbf{G}_y \boldsymbol{\alpha} + \mathbf{G}_y \boldsymbol{\theta} + \mathbf{U}_y \boldsymbol{\psi}_y + \boldsymbol{\epsilon}_y,\end{aligned}$$

where $\mathbf{U}_x \in \mathbb{R}^{n_x \times q}$ and $\mathbf{U}_y \in \mathbb{R}^{n_y \times q}$ are the matrices for q confounders in the samples from IV-to-exposure and IV-to-outcome, respectively, $\boldsymbol{\psi}_x \in \mathbb{R}^{q \times 1}$ and $\boldsymbol{\psi}_y \in \mathbb{R}^{q \times 1}$ are the corresponding vector of coefficients, \mathbf{x}_x and \mathbf{x}_y are exposure traits in two samples,

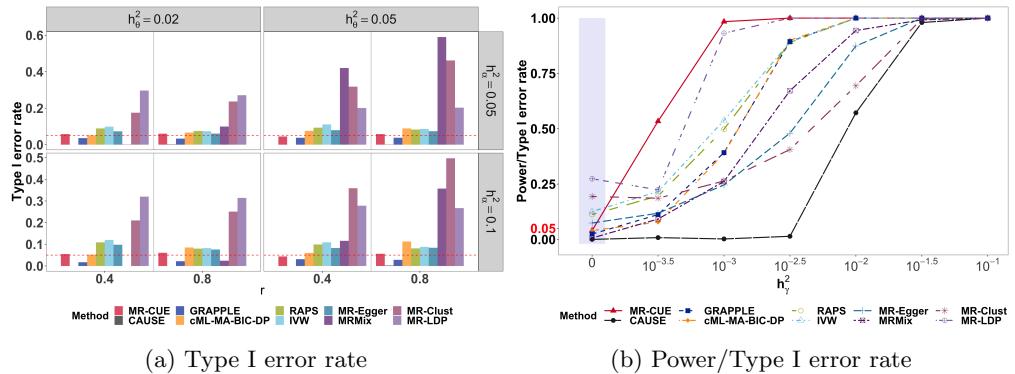
$\epsilon_{x_x} \in \mathbb{R}^{n_x \times 1}$, $\epsilon_{x_y} \in \mathbb{R}^{n_y \times 1}$, and $\epsilon_y \in \mathbb{R}^{n_y \times 1}$ are the random errors, and β_1 is the causal effect of interest. Then we generate summary statistics the same as the scenario of single confounder (see Methods).

3.4 Simulation results for non-linear confounders

For non-linear confounders, we consider the following structural model to generate individual-level data

$$\begin{aligned}\mathbf{x}_x &= \mathbf{G}_x \boldsymbol{\gamma} + \exp(\kappa \mathbf{U}_x \boldsymbol{\psi}_x) + \epsilon_{x_x}, \\ \mathbf{x}_y &= \mathbf{G}_y \boldsymbol{\gamma} + \exp(\kappa \mathbf{U}_y \boldsymbol{\psi}_x) + \epsilon_{x_y}, \\ \mathbf{y} &= \beta_1 \mathbf{x}_y + \mathbf{G}_y \boldsymbol{\alpha} + \mathbf{G}_y \boldsymbol{\theta} + \exp(\kappa \mathbf{U}_y \boldsymbol{\psi}_y) + \epsilon_y\end{aligned}$$

where $\mathbf{U}_x \in \mathbb{R}^{n_x \times q}$ and $\mathbf{U}_y \in \mathbb{R}^{n_y \times q}$ are the matrices for q confounders for exposure and outcome, respectively. We set $\kappa = 0.1$. As shown in Fig. 9, the conclusions on the controls of type I error rates and power comparison are similar to those with linear confounders (Fig. 2 (a) and 2 (c)).



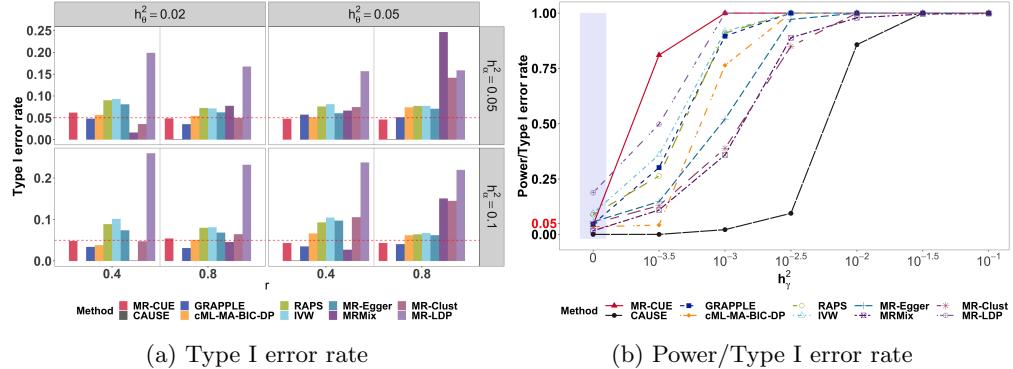
Supplementary Figure 9: Simulation results with non-linear confounders.(a) Type I error rates for MR-CUE and other methods under combinatorial settings for h_θ^2 and h_α^2 with $\rho_{\alpha\gamma} = 0.2$ and $p = 1,000$. (b) Powers for MR-CUE and other methods under the setting: $h_\theta^2 = 0.02$, $h_\alpha^2 = 0.05$, $p = 1,000$, $r = 0.4$ and $\rho_{\alpha\gamma} = 0.2$.

3.5 Simulation results for binary outcome

For binary outcome, we generate data with binary outcome using the following logistic model:

$$\begin{aligned}\mathbf{x}_x &= \mathbf{G}_x \boldsymbol{\gamma} + \mathbf{U}_x \boldsymbol{\psi}_x + \boldsymbol{\epsilon}_{x_x}, \\ \mathbf{x}_y &= \mathbf{G}_y \boldsymbol{\gamma} + \mathbf{U}_y \boldsymbol{\psi}_x + \boldsymbol{\epsilon}_{x_y}, \\ \mathbf{y} &= \text{Bernoulli}(H(\log(1/9) + \beta_1 \mathbf{x}_y + \mathbf{G}_y \boldsymbol{\alpha} + \mathbf{G}_y \boldsymbol{\theta} + \mathbf{U}_y \boldsymbol{\psi}_y))\end{aligned}$$

where $H(t) = 1/(1 + \exp(t))$. The population prevalence is set to be 0.1. We first generate a large population pool of outcomes and sampled 25, 000 cases and 25, 000 controls for the following analysis. As shown in Fig. 10, the conclusions on the controls of type I error rates and power comparison are similar to those with continuous outcome (Fig. 2(a) and 2(c)).

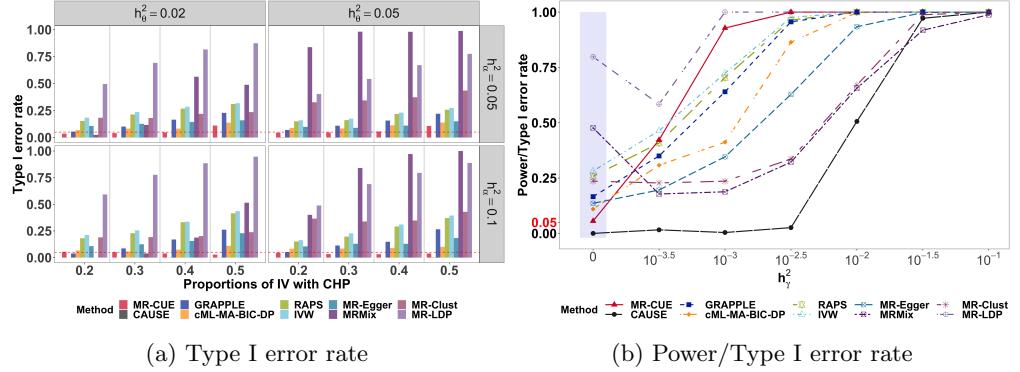


Supplementary Figure 10: Simulation results with binary outcomes. (a) Type I error rates for MR-CUE and other methods under combinatorial settings for h_θ^2 and h_α^2 with $\rho_{\alpha\gamma} = 0.2$ and $p = 1,000$. (b) Powers for MR-CUE and other methods under the setting: $h_\theta^2 = 0.02$, $h_\alpha^2 = 0.05$, $p = 1,000$, $r = 0.4$ and $\rho_{\alpha\gamma} = 0.2$.

3.6 Simulation results for different proportions of IVs with CHP effects

We conduct additional simulations to compare with other methods when the proportions of IV with CHP effect increases. As shown in Fig. 11, when the proportion of IVs with CHP

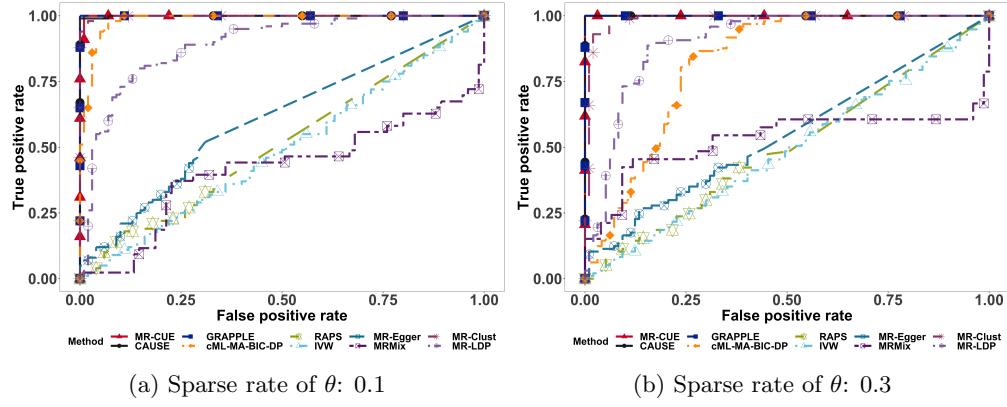
is 0.5, the method CAUSE is conservative and very under-powered. MR-CUE, GRAPPLE and cML-MA-BIT-DP all start to show some loss in the control of type I error rates (with GRAPPLE being the worst and ours the best). MR-CUE is still the most powerful among the methods that have a reasonable control of type I error rates.



Supplementary Figure 11: Simulation results for different proportions of IVs with CHP effects. (a) Type I error rate comparison for MR-CUE and other methods under combinatorial settings for h_θ^2 and h_α^2 with $\rho_{\alpha\gamma} = 0.2$, $r = 0.4$ and $p = 1,000$. (b) Power comparison for MR-CUE and other methods under the setting: $h_\theta^2 = 0.02$, $h_\alpha^2 = 0.05$, $p = 1,000$, $r = 0.4$, $\rho_{\alpha\gamma} = 0.2$ and with the proportions of IVs with CHP effects being 0.4.

3.7 Simulation results for sparse θ in reverse causation analysis

In addition, for the reverse causation analysis, we consider a sparse vector θ , where only a small fraction of SNPs have CHP effects. In Fig. 12, we present methods comparison at different level of UHP sparsity. The conclusions are unchanged.



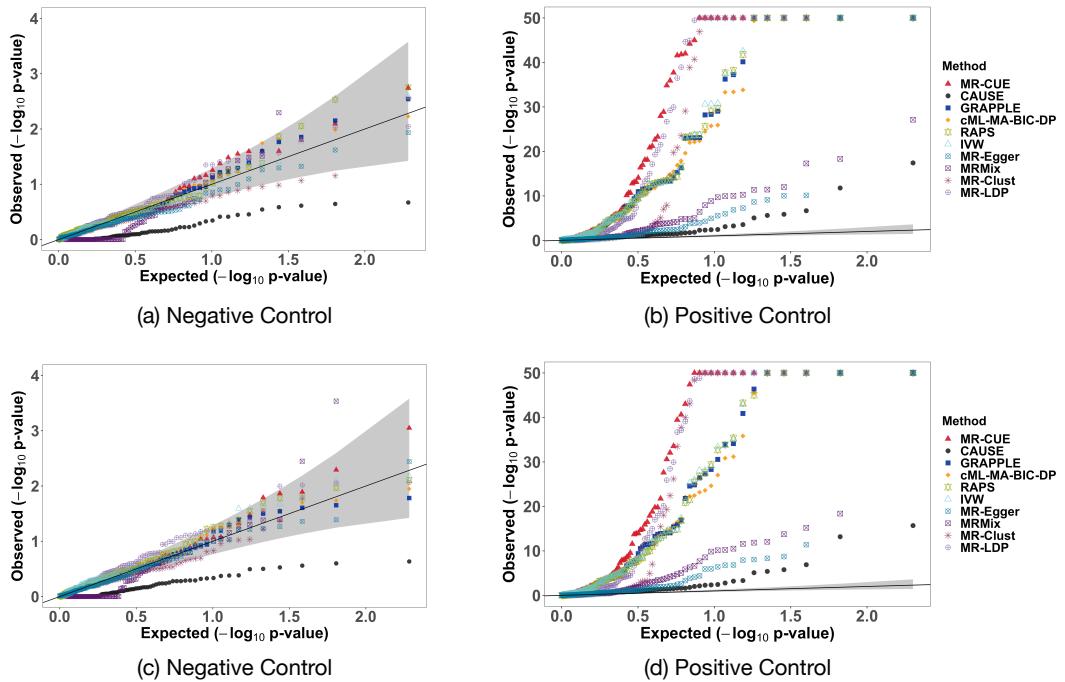
Supplementary Figure 12: The result of reverse causation with different sparse rate of θ .

Supplementary Note 4

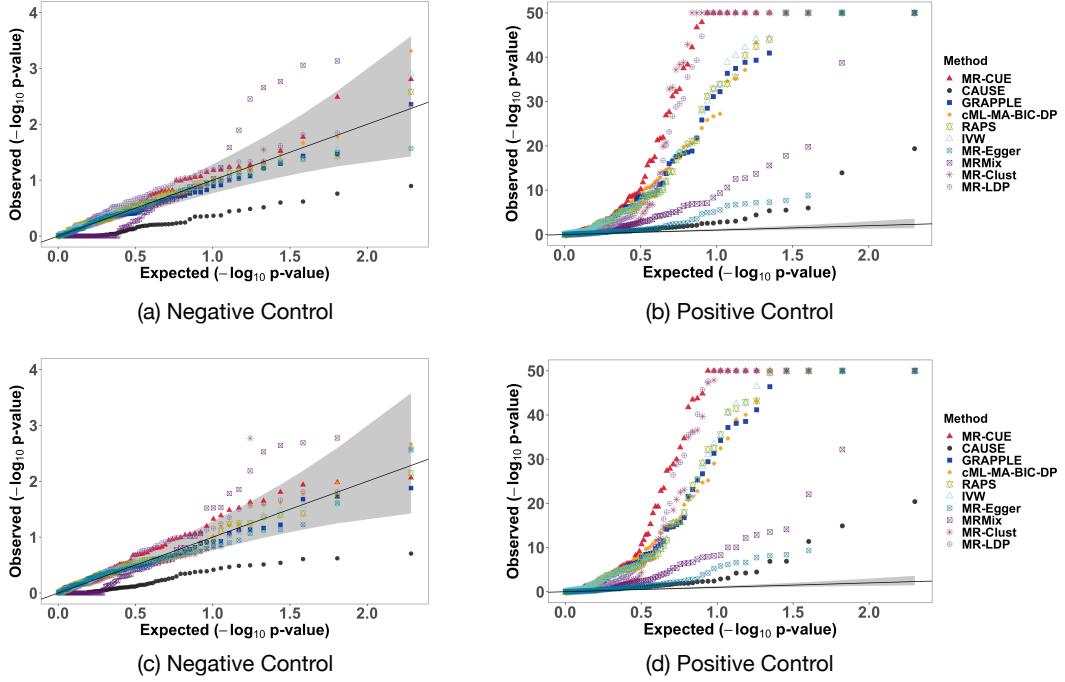
Sensitive Analysis

4.1 Negative and Positive Controls

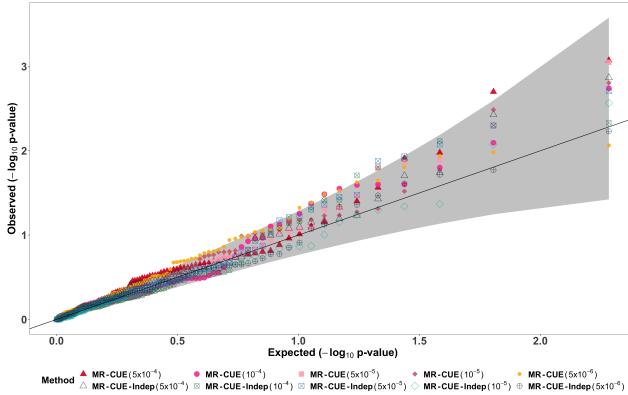
In this section, we adopt different p -value thresholds to select instrument variables for both negative and positive controls, the corresponding quantile-quantile plot are presented in Fig. 13-14. In addition, we display the result of negative control using correlated IVs at different IV selection thresholds, in comparison with the one using only independent IVs (labeled as MR-CUE-Indep), see Fig.15.



Supplementary Figure 13: The result of sensitive analysis with different IV selection thresholds, 10^{-4} (top panel) and 5×10^{-5} (bottom panel), respectively. The grey regions indicate 95% confidence intervals. p -values from all methods are two-sided without adjustment for multiple testing.



Supplementary Figure 14: The result of sensitive analysis with different IV selection thresholds, 10^{-5} (top panel) and 5×10^{-6} (bottom panel), respectively. The grey regions indicate 95% confidence intervals. p -values from all methods are two-sided without adjustment for multiple testing.



Supplementary Figure 15: The result of negative control with correlated and independent SNPs, respectively. Brackets indicate different IV selection thresholds. The grey region indicates a 95% confidence interval. p -values are two-sided without adjustment for multiple testing.

Supplementary Note 5

Real Data Analysis

The estimated causal effects of IL-6 as exposure on 27 outcome traits/disease using MR-CUE and other methods are displayed in Table 3. The estimated causal effects of T2D (European) and T2D (Asian) as outcome using MR-CUE and other methods are presented in Table 4 and Table 5, respectively.

Exposure	MR-CUE	CAUSE	GRAPPLE	cML-MA-BIC-DP	RAPS	IVW	MR-Egger	MRMix	MR-Clust	MR-LDP
SAA	$1.9 \times 10^{-16} \uparrow$	0.309 \downarrow	0.252 \uparrow	0.363 \uparrow	0.560 \uparrow	0.558 \uparrow	0.778 \downarrow	0.903 \downarrow	0.558 \uparrow	$3.8 \times 10^{-38} \uparrow$
HDL	$1.2 \times 10^{-8} \uparrow$	0.999 \uparrow	0.460 \downarrow	0.563 \downarrow	0.804 \downarrow	0.825 \uparrow	0.339 \downarrow	0.616 \uparrow	0.705 \downarrow	$4.6 \times 10^{-7} \uparrow$
LDL	0.003 \uparrow	0.996 \uparrow	0.725 \uparrow	0.744 \uparrow	0.708 \downarrow	0.545 \downarrow	0.764 \downarrow	1.000 \downarrow	0.656 \uparrow	0.086 \uparrow
TG	$5.8 \times 10^{-10} \uparrow$	0.931 \downarrow	0.712 \downarrow	0.868 \downarrow	0.432 \downarrow	0.434 \downarrow	0.955 \downarrow	0.685 \uparrow	0.530 \uparrow	$6.8 \times 10^{-8} \uparrow$
TC	$1.4 \times 10^{-4} \uparrow$	1.000 \uparrow	0.696 \uparrow	0.626 \uparrow	0.711 \downarrow	0.669 \downarrow	0.585 \uparrow	1.000 \downarrow	0.820 \uparrow	0.008 \uparrow
COVID19Sev	$6.2 \times 10^{-55} \downarrow$	0.539 \downarrow	0.165 \downarrow	0.309 \downarrow	0.164 \downarrow	0.176 \downarrow	0.019 \downarrow	0.959 \uparrow	0.176 \downarrow	$9.3 \times 10^{-56} \downarrow$
COVID19Sus	$4.8 \times 10^{-29} \downarrow$	0.998 \downarrow	0.786 \downarrow	0.807 \downarrow	0.908 \downarrow	0.908 \downarrow	0.071 \uparrow	0.772 \uparrow	0.908 \downarrow	$1.2 \times 10^{-34} \downarrow$
AS	$1.8 \times 10^{-30} \downarrow$	0.916 \downarrow	0.387 \downarrow	0.510 \downarrow	0.404 \downarrow	0.414 \downarrow	0.309 \downarrow	0.782 \uparrow	0.414 \downarrow	$0.0 \times 10^0 \downarrow$
AIS	$1.3 \times 10^{-38} \downarrow$	0.810 \downarrow	0.365 \downarrow	0.440 \downarrow	0.438 \downarrow	0.437 \downarrow	0.285 \downarrow	0.792 \uparrow	0.437 \downarrow	$0.0 \times 10^0 \downarrow$
CES	$7.0 \times 10^{-100} \downarrow$	0.930 \downarrow	0.887 \downarrow	0.900 \downarrow	0.653 \downarrow	0.624 \downarrow	0.627 \uparrow	0.442 \downarrow	0.624 \downarrow	$7.0 \times 10^{-211} \downarrow$
SVS	$1.0 \times 10^{-276} \downarrow$	0.714 \downarrow	0.776 \downarrow	0.789 \downarrow	0.715 \downarrow	0.613 \downarrow	0.794 \uparrow	0.458 \downarrow	0.613 \downarrow	$4.9 \times 10^{-324} \downarrow$
T2D	$7.1 \times 10^{-35} \downarrow$	0.440 \downarrow	0.751 \downarrow	0.795 \downarrow	0.673 \downarrow	0.505 \downarrow	0.637 \uparrow	1.000 \downarrow	0.761 \uparrow	0.035 \uparrow
CAD	0.001 \downarrow	0.801 \downarrow	0.179 \downarrow	0.328 \downarrow	0.153 \downarrow	0.155 \downarrow	0.517 \downarrow	0.279 \downarrow	0.242 \downarrow	$1.1 \times 10^{-9} \downarrow$
T1D	$1.5 \times 10^{-53} \downarrow$	0.970 \downarrow	0.585 \downarrow	0.692 \downarrow	0.648 \downarrow	0.679 \downarrow	0.518 \uparrow	0.978 \downarrow	0.804 \uparrow	$3.0 \times 10^{-69} \downarrow$
AD	$8.7 \times 10^{-56} \uparrow$	0.845 \downarrow	0.274 \downarrow	0.469 \downarrow	0.270 \downarrow	0.303 \downarrow	0.212 \downarrow	0.939 \downarrow	0.303 \downarrow	$4.7 \times 10^{-46} \uparrow$
RA	$8.7 \times 10^{-9} \downarrow$	0.991 \downarrow	0.995 \uparrow	0.995 \uparrow	0.862 \uparrow	0.872 \uparrow	0.767 \uparrow	0.851 \downarrow	0.872 \uparrow	$1.2 \times 10^{-117} \downarrow$
CD	$1.3 \times 10^{-206} \downarrow$	0.991 \uparrow	0.300 \downarrow	0.390 \downarrow	0.484 \downarrow	0.330 \downarrow	0.031 \uparrow	0.802 \downarrow	0.330 \downarrow	$2.6 \times 10^{-317} \downarrow$
UC	$1.4 \times 10^{-54} \uparrow$	0.973 \downarrow	0.914 \downarrow	0.993 \downarrow	0.837 \downarrow	0.784 \downarrow	0.666 \uparrow	0.687 \downarrow	0.784 \downarrow	$9.9 \times 10^{-11} \uparrow$
Asthma	0.049 \uparrow	0.984 \uparrow	0.419 \uparrow	0.461 \uparrow	0.257 \uparrow	0.212 \uparrow	0.922 \downarrow	1.000 \downarrow	0.378 \uparrow	0.469 \uparrow
BMD	$3.5 \times 10^{-5} \downarrow$	0.427 \downarrow	0.625 \downarrow	0.742 \downarrow	0.270 \downarrow	0.586 \downarrow	0.182 \downarrow	1.000 \downarrow	0.299 \uparrow	$2.9 \times 10^{-13} \downarrow$
BUN	$4.3 \times 10^{-24} \uparrow$	0.639 \uparrow	0.078 \uparrow	0.141 \uparrow	0.130 \uparrow	0.142 \uparrow	0.218 \uparrow	1.000 \downarrow	0.188 \uparrow	$1.2 \times 10^{-34} \uparrow$
MDD	$1.1 \times 10^{-8} \downarrow$	0.130 \downarrow	0.079 \downarrow	0.245 \downarrow	0.075 \downarrow	0.185 \downarrow	0.002 \downarrow	0.011 \uparrow	0.235 \downarrow	$1.8 \times 10^{-93} \uparrow$
BIP	$8.1 \times 10^{-49} \downarrow$	1.000 \uparrow	0.762 \uparrow	0.852 \uparrow	0.557 \uparrow	0.505 \uparrow	0.290 \downarrow	0.494 \uparrow	0.505 \uparrow	$6.4 \times 10^{-45} \downarrow$
SCZ	0.663 \downarrow	0.883 \uparrow	0.824 \uparrow	0.573 \uparrow	0.567 \uparrow	0.215 \uparrow	0.055 \downarrow	0.193 \uparrow	0.557 \downarrow	$2.8 \times 10^{-6} \uparrow$
eGFR	0.413 \uparrow	0.996 \downarrow	0.133 \uparrow	0.249 \uparrow	0.589 \uparrow	0.732 \uparrow	0.155 \uparrow	1.000 \downarrow	0.203 \uparrow	$2.8 \times 10^{-8} \downarrow$
AF	0.355 \downarrow	0.705 \uparrow	0.828 \uparrow	0.742 \uparrow	0.744 \uparrow	0.360 \uparrow	0.204 \downarrow	0.382 \uparrow	0.493 \downarrow	1.000 \uparrow
Liver	0.793 \downarrow	0.803 \downarrow	0.189 \downarrow	0.353 \downarrow	0.290 \downarrow	0.221 \downarrow	0.592 \downarrow	1.000 \downarrow	0.221 \downarrow	$4.7 \times 10^{-4} \downarrow$

Supplementary Table 3: Summary results for evaluating the causal effects of IL-6 on 27 outcome traits/disease. Columns 2-11 listed the p -values for each MR method, p -values are marked in bold if $p < 0.05$. Arrows indicate the signs of the corresponding effect estimates. p -values from all methods are two-sided without adjustment for multiple testing.

Exposure	MR-CUE	CAUSE	GRAPPLE	cML-MA-BIC-DP	RAPS	IVW	MR-Egger	MRMix	MR-Clust	MR-LDP
HIP	$3.4 \times 10^{-14} \uparrow$	0.058 \uparrow	0.002 \downarrow	0.001 \downarrow	$1.9 \times 10^{-4} \uparrow$	$3.0 \times 10^{-5} \uparrow$	0.321 \uparrow	0.271 \downarrow	0.809 \downarrow	0.204 \downarrow
WC	$7.6 \times 10^{-41} \downarrow$	$7.5 \times 10^{-5} \uparrow$	$3.7 \times 10^{-42} \uparrow$	$4.5 \times 10^{-62} \uparrow$	0.000 \uparrow	$5.3 \times 10^{-18} \uparrow$	0.709 \uparrow	$2.2 \times 10^{-47} \uparrow$	0.000 \uparrow	$1.1 \times 10^{-46} \uparrow$
WHR	$6.1 \times 10^{-14} \uparrow$	$2.1 \times 10^{-4} \uparrow$	$2.6 \times 10^{-24} \uparrow$	$1.8 \times 10^{-29} \uparrow$	$3.7 \times 10^{-14} \uparrow$	$1.7 \times 10^{-27} \uparrow$	0.286 \uparrow	$6.6 \times 10^{-7} \uparrow$	0.067 \downarrow	$4.9 \times 10^{-35} \uparrow$
BMI	$8.8 \times 10^{-99} \downarrow$	$8.6 \times 10^{-4} \uparrow$	$1.2 \times 10^{-51} \uparrow$	$6.0 \times 10^{-115} \uparrow$	0.000 \uparrow	$1.8 \times 10^{-25} \uparrow$	0.023 \uparrow	$1.0 \times 10^{-65} \uparrow$	0.000 \uparrow	$2.4 \times 10^{-112} \uparrow$
BFP	$2.1 \times 10^{-10} \uparrow$	0.291 \uparrow	0.179 \uparrow	0.530 \uparrow	0.117 \uparrow	0.336 \uparrow	0.727 \downarrow	$2.9 \times 10^{-50} \uparrow$	0.977 \uparrow	$1.4 \times 10^{-30} \uparrow$
BL	$1.1 \times 10^{-11} \downarrow$	0.936 \downarrow	0.104 \downarrow	0.135 \downarrow	0.102 \downarrow	0.112 \downarrow	0.686 \downarrow	0.320 \downarrow	0.172 \uparrow	$2.3 \times 10^{-23} \downarrow$
BW	$2.9 \times 10^{-4} \downarrow$	0.200 \downarrow	$2.0 \times 10^{-4} \downarrow$	$7.4 \times 10^{-4} \downarrow$	$5.0 \times 10^{-4} \downarrow$	$2.0 \times 10^{-6} \downarrow$	0.449 \downarrow	$5.7 \times 10^{-6} \downarrow$	0.015 \uparrow	$1.6 \times 10^{-56} \downarrow$
HR	$1.2 \times 10^{-14} \uparrow$	0.702 \downarrow	0.392 \downarrow	0.417 \downarrow	0.508 \downarrow	0.516 \downarrow	0.865 \downarrow	1.000 \downarrow	0.564 \downarrow	0.009 \downarrow
HF	$2.3 \times 10^{-202} \uparrow$	0.212 \uparrow	$1.7 \times 10^{-22} \uparrow$	$1.5 \times 10^{-22} \uparrow$	0.000 \uparrow	$7.8 \times 10^{-29} \uparrow$	0.512 \uparrow	$1.2 \times 10^{-4} \uparrow$	0.298 \downarrow	0.000 \uparrow
HDL-C	$1.0 \times 10^{-13} \downarrow$	0.027 \downarrow	$8.6 \times 10^{-15} \downarrow$	$6.9 \times 10^{-13} \downarrow$	$2.2 \times 10^{-16} \downarrow$	$1.8 \times 10^{-18} \downarrow$	0.497 \downarrow	0.114 \downarrow	0.034 \uparrow	$5.6 \times 10^{-73} \downarrow$
TG	0.950 \uparrow	0.022 \uparrow	$1.0 \times 10^{-13} \uparrow$	$5.8 \times 10^{-17} \uparrow$	$3.8 \times 10^{-14} \uparrow$	$1.3 \times 10^{-18} \uparrow$	0.320 \downarrow	0.005 \uparrow	0.047 \downarrow	$1.9 \times 10^{-24} \uparrow$
SBP	0.000 \uparrow	0.011 \uparrow	$2.1 \times 10^{-21} \uparrow$	$3.0 \times 10^{-21} \uparrow$	0.000 \uparrow	$3.4 \times 10^{-23} \uparrow$	0.015 \uparrow	$7.2 \times 10^{-4} \uparrow$	0.002 \downarrow	0.000 \uparrow
DBP	0.000 \uparrow	0.038 \uparrow	$4.1 \times 10^{-9} \uparrow$	$6.0 \times 10^{-8} \uparrow$	$8.9 \times 10^{-10} \uparrow$	$6.2 \times 10^{-10} \uparrow$	0.041 \uparrow	0.166 \uparrow	0.552 \downarrow	0.000 \uparrow
PP	0.000 \uparrow	0.007 \uparrow	$8.3 \times 10^{-23} \uparrow$	$5.2 \times 10^{-24} \uparrow$	0.000 \uparrow	$4.7 \times 10^{-29} \uparrow$	0.003 \uparrow	$1.3 \times 10^{-9} \uparrow$	0.186 \downarrow	0.000 \uparrow
AIS	$4.4 \times 10^{-8} \uparrow$	0.616 \uparrow	$4.9 \times 10^{-11} \uparrow$	$3.1 \times 10^{-10} \uparrow$	$7.1 \times 10^{-13} \uparrow$	$2.0 \times 10^{-14} \uparrow$	0.102 \uparrow	0.008 \uparrow	0.072 \downarrow	$8.1 \times 10^{-299} \uparrow$
MDD	0.000 \uparrow	0.146 \uparrow	$2.2 \times 10^{-13} \uparrow$	$1.1 \times 10^{-14} \uparrow$	$1.5 \times 10^{-11} \uparrow$	$5.6 \times 10^{-14} \uparrow$	0.119 \uparrow	$4.9 \times 10^{-4} \uparrow$	0.861 \uparrow	0.000 \uparrow
UA	$1.0 \times 10^{-13} \downarrow$	0.956 \uparrow	0.004 \uparrow	0.002 \uparrow	0.001 \uparrow	$4.3 \times 10^{-8} \uparrow$	0.475 \downarrow	0.111 \uparrow	0.489 \downarrow	$6.0 \times 10^{-18} \uparrow$
RBC	$3.6 \times 10^{-46} \downarrow$	0.913 \uparrow	0.045 \uparrow	0.004 \uparrow	0.151 \uparrow	0.001 \uparrow	0.103 \downarrow	0.017 \uparrow	0.007 \downarrow	$2.5 \times 10^{-134} \downarrow$
WBC	$1.2 \times 10^{-93} \uparrow$	0.931 \uparrow	$1.1 \times 10^{-5} \uparrow$	$6.7 \times 10^{-6} \uparrow$	$3.5 \times 10^{-5} \uparrow$	$2.6 \times 10^{-8} \uparrow$	0.845 \downarrow	0.016 \uparrow	0.034 \downarrow	$9.7 \times 10^{-172} \uparrow$
PLT	$6.8 \times 10^{-34} \downarrow$	0.995 \downarrow	0.828 \downarrow	0.892 \uparrow	0.392 \downarrow	0.510 \downarrow	0.488 \downarrow	0.360 \uparrow	0.862 \uparrow	$3.4 \times 10^{-28} \downarrow$
Glu	0.196 \downarrow	0.851 \uparrow	0.029 \uparrow	0.103 \uparrow	0.021 \uparrow	0.009 \uparrow	0.805 \uparrow	0.680 \downarrow	0.629 \uparrow	$5.6 \times 10^{-17} \uparrow$
Gly	$2.3 \times 10^{-8} \uparrow$	0.699 \uparrow	0.884 \uparrow	0.788 \uparrow	0.979 \downarrow	0.928 \uparrow	0.986 \downarrow	0.412 \uparrow	0.748 \downarrow	<

Exposure	MR-CUE	CAUSE	GRAPPLE	cML-MA-BIC-DP	RAPS	IVW	MR-Egger	MRMix	MR-Clust	MR-LDP
BMI	$2.1 \times 10^{-27} \uparrow$	0.076 \uparrow	$3.7 \times 10^{-8} \uparrow$	$2.3 \times 10^{-16} \uparrow$	0.066 \uparrow	0.089 \uparrow	0.050 \uparrow	$2.9 \times 10^{-11} \uparrow$	$3.3 \times 10^{-6} \downarrow$	$8.2 \times 10^{-51} \uparrow$
HF	0.004 \uparrow	0.992 \uparrow	0.001 \uparrow	0.005 \uparrow	$8.6 \times 10^{-4} \uparrow$	0.001 \uparrow	0.437 \uparrow	0.103 \uparrow	0.165 \uparrow	0.213 \uparrow
HDL-C	$3.8 \times 10^{-11} \downarrow$	1.000 \downarrow	0.038 \downarrow	0.009 \downarrow	0.037 \downarrow	0.007 \downarrow	0.042 \uparrow	0.055 \downarrow	0.567 \uparrow	$4.3 \times 10^{-4} \downarrow$
TG	$1.8 \times 10^{-36} \uparrow$	0.586 \uparrow	$1.0 \times 10^{-4} \uparrow$	$5.9 \times 10^{-5} \uparrow$	0.006 \uparrow	0.012 \uparrow	0.006 \downarrow	0.002 \uparrow	0.142 \downarrow	0.001 \uparrow
SBP	0.209 \uparrow	0.145 \uparrow	$2.1 \times 10^{-6} \uparrow$	$3.4 \times 10^{-6} \uparrow$	$1.3 \times 10^{-5} \uparrow$	$2.1 \times 10^{-5} \uparrow$	0.653 \uparrow	0.003 \uparrow	0.066 \uparrow	$6.6 \times 10^{-9} \uparrow$
DBP	0.003 \uparrow	1.000 \downarrow	0.249 \uparrow	0.270 \uparrow	0.814 \downarrow	0.591 \downarrow	0.807 \downarrow	0.274 \uparrow	0.708 \downarrow	0.188 \uparrow
PP	$7.6 \times 10^{-23} \uparrow$	0.837 \uparrow	0.018 \uparrow	0.025 \uparrow	0.009 \uparrow	0.010 \uparrow	0.435 \uparrow	0.211 \uparrow	0.529 \uparrow	$1.6 \times 10^{-27} \uparrow$
AIS	$1.2 \times 10^{-5} \uparrow$	0.331 \uparrow	$3.1 \times 10^{-4} \uparrow$	$5.6 \times 10^{-4} \uparrow$	$4.8 \times 10^{-4} \uparrow$	$1.7 \times 10^{-4} \uparrow$	0.899 \downarrow	0.026 \uparrow	0.629 \uparrow	$8.7 \times 10^{-11} \uparrow$
RBC	$4.0 \times 10^{-23} \uparrow$	0.184 \uparrow	$1.3 \times 10^{-7} \uparrow$	$5.5 \times 10^{-6} \uparrow$	$4.2 \times 10^{-6} \uparrow$	$9.6 \times 10^{-7} \uparrow$	0.740 \uparrow	$2.6 \times 10^{-5} \uparrow$	0.462 \uparrow	$1.4 \times 10^{-45} \downarrow$
WBC	$2.5 \times 10^{-36} \uparrow$	0.510 \uparrow	$2.0 \times 10^{-4} \uparrow$	$2.5 \times 10^{-4} \uparrow$	$2.5 \times 10^{-5} \uparrow$	$6.0 \times 10^{-5} \uparrow$	0.676 \uparrow	0.010 \uparrow	0.424 \uparrow	$4.3 \times 10^{-6} \uparrow$
PLT	0.603 \uparrow	0.993 \downarrow	0.071 \downarrow	0.040 \downarrow	0.022 \downarrow	0.038 \downarrow	0.421 \downarrow	0.142 \downarrow	0.145 \downarrow	$1.1 \times 10^{-13} \uparrow$
Glu	$4.3 \times 10^{-5} \downarrow$	0.930 \downarrow	0.417 \downarrow	0.489 \downarrow	0.307 \downarrow	0.337 \downarrow	0.462 \downarrow	1.000 \downarrow	0.529 \downarrow	$8.2 \times 10^{-32} \uparrow$
Gly	$3.3 \times 10^{-13} \uparrow$	1.000 \uparrow	0.462 \downarrow	0.508 \downarrow	0.809 \downarrow	0.906 \downarrow	0.258 \downarrow	0.705 \uparrow	0.679 \downarrow	$6.2 \times 10^{-68} \downarrow$
HbA1c	$4.8 \times 10^{-4} \uparrow$	0.990 \uparrow	0.403 \uparrow	0.648 \uparrow	0.145 \uparrow	0.132 \uparrow	0.215 \uparrow	0.479 \downarrow	0.408 \downarrow	1.000 \uparrow

Supplementary Table 5: Summary results for evaluating the causal effects of 14 exposures on T2D risk in the East Asian population. Columns 2-11 listed the p -values for each MR method, p -values are marked in bold if $p < 0.05$. Arrows indicate the signs of the corresponding effect estimates. p -values from all methods are two-sided without adjustment for multiple testing.

Supplementary References

- [1] Morrison, J., Knoblauch, N., Marcus, J. H., Stephens, M. & He, X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nature genetics* **52**, 740–747 (2020).
- [2] Xue, H., Shen, X. & Pan, W. Constrained maximum likelihood-based mendelian randomization robust to both correlated and uncorrelated pleiotropic effects. *The American Journal of Human Genetics* **108**, 1251–1269 (2021).
- [3] Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic epidemiology* **40**, 304–314 (2016).
- [4] Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nature genetics* **50**, 693–698 (2018).